(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 March 2003 (20.03.2003)

PCT

(10) International Publication Number
WO 03/023013 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/29001

(22) International Filing Date:
13 September 2002 (13.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/322,511	13 September 2001 (13.09.2001)	US
10/243,552	12 September 2002 (12.09.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/322,511 (CIP)
Filed on	13 September 2001 (13.09.2001)
US	10/243,552 (CIP)
Filed on	12 September 2002 (12.09.2002)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YANG, Yonghong** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **WANG, Zhiwei** [CN/US]; 836 Alturas Avenue #B36, Sunnyvale, CA 94085 (US). **WENG, Gezhi** [US/US]; 415 Moraga Avenue, Piedmont, CA 94611 (US). **MA, Yunqing** [CN/US]; 280 W. California Avenue #206, Sunnyvale, CA 94086 (US).

(74) Agents: **RIN-LAURES, Li-Hsien** et al.; Hyseq, Inc., 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/023013 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Application Serial No. 60/322,511 filed September 13, 2001 entitled "Novel Nucleic Acids and Polypeptides", Attorney Docket No. 807, which in turn is a continuation-in-part application of PCT Application Serial No. PCT/US00/35017 filed December 22, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP3A/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/552,317 filed April 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/488,725 filed January 21, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784; PCT Application Serial No. PCT/US01/02623 filed January 25, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/491,404 filed January 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785; PCT Application Serial No. PCT/US01/03800 filed February 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/560,875 filed April 27, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/496,914 filed February 03, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787; PCT Application Serial No. PCT/US01/04927 filed February 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,409 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/515,126 filed February 28, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788; PCT Application Serial No. PCT/US01/04941 filed March 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/574,454 filed May 19, 2000 entitled "Novel Contigs Obtained from Various

Libraries", Attorney Docket No. 789CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/519,705 filed March 07, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789; PCT Application Serial No. PCT/US01/08631 filed March 30, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/649,167 filed August 23, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/540,217 filed March 31, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790; PCT Application Serial No. PCT/US01/08656 filed April 18, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/770,160 filed January 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP, which is in turn a continuation-in-part application of U.S. Application Serial No. 09/552,929 filed April 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791; and PCT Application Serial No. PCT/US01/14827 filed May16, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,408 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 792; all of which are incorporated herein by reference in their entirety.

2. BACKGROUND OF THE INVENTION

2.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, circulating soluble factors, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression

cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-336, or 673-873 and are provided in

the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases or unknown. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-336, or 673-873 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-336, or 673-873. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-336, or 673-873 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-336, or 673-873. The sequence information can be a segment of any one of SEQ ID NO: 1-336, or 673-873 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-336, or 673-873.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-336, or 673-873 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the

nucleic acid sequences of SEQ ID NO: 1-336, or 673-873 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-336, or 673-873; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-336, or 673-873; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-336, or 673-873. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-336, or 673-873; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in SEQ ID NO: 1-336, or 673-873; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homologue (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in SEQ ID NO: 337-672, or 874-1074, or Tables 3, 4A, 4B, 5, 6, or 8.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-336, or 673-873; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such processes is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for

treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can affect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Tables 2A and 2B); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Tables 4A and 4B). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only certain portion(s) of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded

molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G, or T (U) or unknown. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is

capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NO: 1-336, or 673-873.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-336, or 673-873. The sequence information can be a segment of any one of SEQ ID NO: 1-336, or 673-873 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-336, or 673-873, or those segments identified in Tables 3, 4A, 4B, 5, 6, or 8. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three

billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids.

Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full-length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be

reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or

elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of

those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a new stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal

integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-336, or 673-873; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 1-336, or 673-873; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-336, or 673-873. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1-336, or 673-873; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homologue of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 337-672, or 874-1074 (for example, as set forth in Tables 3, 4A, 4B, 5, 6, or 8). Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable

immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-336, or 673-873 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-336, or 673-873 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-336, or 673-873 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide

sequences of SEQ ID NO: 1-336, or 673-873, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-336, or 673-873, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-336, or 673-873 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology results for the nucleic acids of the present invention, including SEQ ID NO: 1-336, or 673-873 can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST (Basic Local Alignment Search Tool) program is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using FASTXY algorithm may be performed.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be

prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA

fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention could be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-336, or 673-873, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient

restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-336, or 673-873 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-336, or 673-873 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example: Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse

metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., Nat. Biotech 17, 870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-336, or 673-873, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1-336, or 673-873 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-336, or 673-873 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-336, or 673-873, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more

preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the

case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1-336, or 673-873). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease

activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA

portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory

Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, and regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 337-672, or 874-1074 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-336, or 673-873 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NO: 1-336, or 673-873 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 337-672, or 874-1074 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 337-672, or 874-1074 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 337-672, or 874-1074.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as

immunoglobulins for many purposes, including increasing the valency of protein binding sites. Fragments are also identified in Tables 3, 4A, 4B, 5, 6, or 8.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The predicted signal sequence is set forth in Table 6. The mature form of such protein may be obtained and confirmed by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell and sequencing of the cleaved product. One of skill in the art will recognize that the actual cleavage site may be different than that predicted in Table 6. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed (See, e.g., Sakal et al., Prep. Biochem. Biotechnol. (2000), 30(2), pp. 107-23, incorporated herein by reference).

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may

be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 337-672, or 874-1074.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.,* as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.,* U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide

a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent.. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), Pfam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), the GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95,

13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) incorporated herein by reference).

Polypeptide sequences were examined by a proprietary algorithm, SeqLoc that separates the proteins into three sets of locales: intracellular, membrane, or secreted. This prediction is based upon three characteristics of each polypeptide, including percentage of cysteine residues, Kyte-Doolittle scores for the first 20 amino acids of each protein, and Kyte-Doolittle scores to calculate the longest hydrophobic stretch of the said protein. Values of predicted proteins are compared against the values from a set of 592 proteins of known cellular localization from the Swissprot database (http://www.expasy.ch/sprot). Predictions are based upon the maximum likelihood estimation.

Pesence of transmembrane region(s) was detected using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

. The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be

modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.,* inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are

deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by referenee.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA

sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid

preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors.

The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies

would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the

invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast

activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from

chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and

disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of

an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro

antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to

tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis

Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention

(including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987)

Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening

assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.,* increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The responses of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an

inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include

but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival

or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution,

change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that

hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

5

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode
10 of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of
15 polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution,
20 dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

25 **4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION**

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be
30 administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other

materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti- inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound

sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in

fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical

composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired,

disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such

as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

. A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable

matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides,

diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet

derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be

estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 337-672, or 874-1074, or Tables 3, 4A, 4B, 5, 6, or 8, or 9, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine

binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known

in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256, 495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107, 220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as

a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.13.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321, 522-525 (1986); Riechmann et al., Nature, 332, 323-327 (1988); Verhoeyen et al., Science, 239, 1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion

of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2, 593-596 (1992)).

4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80, 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227, 381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368, 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13, 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains

in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246, 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305, 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10, 3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121, 210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229, 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175, 217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical

coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5), 1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90, 6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152, 5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147, 60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176, 1191-1195 (1992) and Shopes, J. Immunol., 148, 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53, 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3, 219-230 (1989).

4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the

presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-336, or 673-873 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-336, or 673-873 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the

present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,

but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix-see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 15241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-336, or 673-873, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"

when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Okano, J. Neurochem. 56, 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-336, or 673-873. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NO: 1-336, or 673-873 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well-known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal

map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

4.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6), 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8), 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridgeheads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins

the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/μl) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-$MeIm_7$), is then added to a final concentration of 10 mM 1-$MeIm_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 μl/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-$MeIm_7$, is made fresh and 25 μl added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995), 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res., 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1), 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Nat'l. Acad. Sci., USA 91(11), 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24), 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 μg instead of 2-5 μg); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient.

Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5.0 EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences.

5.2 EXAMPLE 2

Assemblage of Novel Contigs

The contigs of the present invention, designated as SEQ ID NO: 673-873 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene, and exons from public domain genomic sequences predicated by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above databases that would extend the assemblage. Further, inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

5.3 EXAMPLE 3

Novel Nucleic Acids

The novel nucleic acids of the present invention SEQ ID NO: 1-336 were assembled from Hyseq 's proprietary EST sequences as described in Example 1 and human genome sequences that are available from the public databases (http://www.ncbi.nlm.nih.gov/). Exons were predicted from human genome sequences using GenScan (http://genes.mit.edu/GENSCANinfo.html); HMMgene (http://www.cbs.dtu.dk/services/HMMgene/hmmgene1_1.html); and GenMark.hmm (http://genemark.biology.gatech.edu/GeneMark/whmm_info.html). The Hyseq proprietary EST sequences and the predicted exons were assembled based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%. Then, the predicted genes were analyzed using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) for presence of a signal peptide. These sequences were further analyzed for presence of transmembrane region(s) using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

Table 1 shows the various tissue sources of SEQ ID NO: 1-336.

The homologs for polypeptides SEQ ID NO: 337-672, that correspond to nucleotide sequences SEQ ID NO: 1-336 were obtained by a BLASTP search against Genpept release 124 and Geneseq (Derwent) release 200117 and against Genpept release 129 and Geneseq (Derwent) release (July 18, 2002). The results showing homologues for SEQ ID NO: 337-672 from Genpept 124 are shown in Table 2A. The results showing homologues for SEQ ID NO: 337-672 from Genpept 129 are shown in Table 2B.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6, 219-235 (1999), http://motif.stanford.edu/ematrix-search/ herein incorporated by reference), all the polypeptide sequences were examined to determine whether they had identifiable signature regions. Scoring matrices of the eMatrix software package are derived from the BLOCKS, PRINTS, PFAM, PRODOM, and DOMO databases. Table 3 shows the accession number of the homologous eMatrix signature found in the indicated polypeptide sequence, its description, and the results obtained which include accession number subtype; raw score; p-value; and the position of signature in amino acid sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4A shows the name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence as described in United States priority application serial number 60/322,511, filed September 13, 2001, herein incorporated by reference in its entirety. Table 4B shows the name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence using Pfam version 7.2. Further description of the Pfam models can be found at http://pfam.wustl.edu/.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ ID NO: 1-336 (i.e. SEQ ID NO: 337-672). Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by Altschul et al, (Nucl. Acids. Res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA,) which is an automated sequence and structure searching procedure (http://www.msi.com/), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with

the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows: "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as annotated by the PDB files (http:/www.rcsb.org/PDB/); start and end amino acid position of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score is produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David Eisenberg's laboratory (US patent no. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:13597-12502. The verify score produced by GeneAtlas normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

Verify score (normalized) = (raw score – 1/2 high score)/(1/2 high score)

The PFM score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potentials (MFP). As given in table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good model. Similarly, a PMF score between 0 to 1.0, with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may also be determined by one of skill in the art based all the information in Table 5 taken in totality.

Table 6 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean

S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences.

Table 7 correlates each of SEQ ID NO: 1-336 to a specific chromosomal location.

Table 8 shows the number of transmembrane regions, their location(s), and TMPred score obtained, for each of the SEQ ID NO: 337-672 that had a TMPred score of 800 or greater, using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

Table 9 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-336, their corresponding polypeptide sequences SEQ ID NO: 337-672, their corresponding priority nucleotide sequences SEQ ID NO: 673-873, their corresponding priority polypeptide sequences SEQ ID NO: 874-1074, and the US serial number of the priority application in which the sequence was filed.

Table 10 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-336, the novel polypeptide sequences SEQ ID NO: 337-672, and the corresponding SEQ ID NO in which the sequence was filed in priority US application bearing serial number 60/322,511, filed September 13, 2001.

Table 1

115

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
adrenal gland	Clontech	ADR002	25 29 72 79 176 220 233 246 270 285 287 298-299 306
adult bladder	Invitrogen	BLD001	317 321 331
adult brain	Clontech	ABR001	29 176 207 211 288 306 332
adult brain	Clontech	ABR006	3 35 49 59 62 69 71 73-74 96 98-99 101-102 108 111 129 161-166 168-172 185-186 196 201 211 283 316 321
adult brain	Clontech	ABR008	13 15-20 24 43 57-59 63 65 74 83 96 105 108 125 129 135 158 162 193 201 207 218-219 248-249 261 263 274 278 285 289-292 294-297 312-313 316-317 321-324 330 332
adult brain	GIBCO	AB3001	78 123 134 182 265 318
adult brain	GIBCO	ABD003	9-10 24 64-65 102 108 119 145 149 182 211 253 263 265 296 318
adult brain	Invitrogen	ABR014	207 248 318
adult brain	Invitrogen	ABR015	296 318
adult brain	Invitrogen	ABR016	228 296
adult brain	Invitrogen	ABT004	4 72 81 193 196 207 274-275 295 306-307
adult cervix	BioChain	CVX001	22 24 69 72 75 83 102 111 149 185 265 278 287 291 303 306 318 331
adult colon	Invitrogen	CLN001	144 175 182-183 245
adult heart	GIBCO	AHR001	13 16 23-24 31 34 57 65 71 96 185 195 236 257 265-266 277 306-307 318 321 326 336
adult kidney	GIBCO	AKD001	12 45 57 65 72 83 106 149 175 178 182 202 205 207 234 265 278 285-287 300 307 318 326
adult kidney	Invitrogen	AKT002	10 25 29 65 72 82-83 119 142 166 175-176 205 211 224 236 270 272 278 287 326
adult liver	Clontech	ALV003	222
adult liver	Invitrogen	ALV002	21 44 46 49 65 72 83 130 149 182 197 234 240 277 308
adult lung	GIBCO	ALG001	10 234
adult ovary	Invitrogen	AOV001	4 10 25 29 37 46 65 70 72 124 129 131 142 149-150 154 156 176 178 188 206-207 211 219 230 245 265 270 276-278 283-284 287 318

Table 1

116

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			321
adult placenta	Clontech	APL001	129 194 318
adult spleen	Clontech	SPLc01	13 46 69 143 152 179 300 317
adult spleen	GIBCO	ASP001	46 79 182 207 317-318
adult testis	GIBCO	ATS001	167 236 269 318
bone marrow	Clontech	BMD001	6 56 69 78 129 267 277-279 282 304 315 318
bone marrow	GF	BMD002	1 6 13 39-40 48-49 60 91-94 102-103 136 143 234 236 277-279 285 291 297 330 336
cultured preadipocytes	Stratagene	ADP001	41-43 182 275 277 314 326
endothelial cells	Stratagene	EDT001	10 25 49 72 105 110 130 206-207 236 270 272 277-278 310 312 318-319
fetal brain	Clontech	FBR001	164 185
fetal brain	Clontech	FBR004	184-185
fetal brain	Clontech	FBR006	6 13 15 46 49 61-63 72 83 96 100 102 107-108 110 135 146 162 186-188 190-194 203 207 219 224 236 262 274 284 291 303 316 323 331-334
fetal brain	GIBCO	HFB001	43 47 64 71-72 100 112 130 154 178 182 189 236 245 265 277 293 296 318
fetal brain	Invitrogen	FBT002	47 49 72 207 276 287 291
fetal heart	Invitrogen	FHR001	6 17 19 31 49 108 113-114 126-128 142 177 182 201 207 243 279 284 300 316
fetal kidney	Clontech	FKD001	5 25 270
fetal kidney	Clontech	FKD002	13 70 96 115-116 136 138 164 193 201 205 292 317 324
fetal liver	Clontech	FLV002	17 222 292 326
fetal liver	Clontech	FLV004	1 49 96 117 137-138 186 222 236 239 263 303 324
fetal liver	Invitrogen	FLV001	72 207 233 273 310 331
fetal liver-spleen	Columbia University	FLS001	10 13 25 31-32 43 67 69 80 83-85 89 111 123-124 129-130 132 150 186 201 207 222 245 254 257 270-272 277-279 281 287 302-303 306 314 318
fetal liver-spleen	Columbia University	FLS002	12 25 32-33 35-36 43 46 48-49 65 69 72 77 83-84 124 129-131 142 148 174 198 206-207 222

Table 1

117

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			229 231 239 244-245 254 257 270 272 277 279 302 314 318 327
fetal liver-spleen	Columbia University	FLS003	31-32 46 70 87-88 90 123 132-133 142 222
fetal lung	Clontech	FLG001	4 112
fetal lung	Invitrogen	FLG003	30 335-336
fetal muscle	Invitrogen	FMS001	72 182 207 236 310
fetal muscle	Invitrogen	FMS002	46 57 104-106 139-140 236 318
fetal skin	Invitrogen	FSK001	2 17 30 43 65-66 125 166 172 182 207 236 255 300 314
fetal skin	Invitrogen	FSK002	8 13 17 19 43 49 57 75-76 108 118-120 141-144 146-148 177 186 236 255 297 324 326 331
fibroblast	Stratagene	LFB001	318
induced neuron-cells	Stratagene	NTD001	185 236 283
infant brain	Columbia University	IB2002	8-10 43 65 71 152 157 162 182 189 211 248 263 280 285 309 328
infant brain	Columbia University	IB2003	21 47 153-155 157-158 169 185 188 211 268 278 306 309 314
infant brain	Columbia University	IBM002	182 331
infant brain	Columbia University	IBS001	72 211 268 278
leukocyte	Clontech	LUC003	149 221
leukocyte	GIBCO	LUC001	1 11 49 68 96 149 176 182 189 223 232 236 245 273 278-279 287 291 314 318 325-326 331
lung	318		
lung tumor	Invitrogen	LGT002	21 24-28 46 49 72 89 175 193 200 205-207 223 236 241 245 256 277 292 294 307 310-312 314 318 326 329
lymph node	Clontech	ALN001	169 263 318
lymphocytes	ATCC	LPC001	19 37 49 68 77 123 143 149-151 189 207 260 263 278 325 330-331
macrophage	Invitrogen	HMP001	202 251
mammary gland	Invitrogen	MMG001	21 67 83 125 131 174 182 193 205 211 223 234 238 263 265 277 287 300-302 313-314 318 321
melanoma from-cell-line-ATCC-#CRL-1424	Clontech	MEL004	7 203-204 207 287
*Mixture of 16 tissues - mRNA	Various Vendors	CGd010	227
*Mixture of 16 tissues - mRNA	Various Vendors	CGd011	124 182 302
*Mixture of 16 tissues - mRNA	Various Vendors	CGd012	14 17 32 42 107 119 124 146 162 206 228-230

Table 1

118

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			236 239-244 246-247 256 265 272 302
*Mixture of 16 tissues - mRNA	Various Vendors	CGd013	18 229-230 311
*Mixture of 16 tissues - mRNA	Various Vendors	CGd015	32 231 252-253 272 307 310
*Mixture of 16 tissues - mRNA	Various Vendors	CGd016	25 39 161 178 236 248-250
neuronal cells	Stratagene	NTU001	174 207
pituitary gland	Clontech	PIT004	180 207 236 283 291
placenta	Clontech	PLA003	19 96 119 121-122 143 148 161 255
placenta	Invitrogen	APL002	207 314
prostate	Clontech	PRT001	131 136 149 181 292 307 327
rectum .	Invitrogen	REC001	3 38 131 166 182 324
retinoic acid-induced-neuronal-cells	Stratagene	NTR001	111 130 173 207 287
salivary gland	Clontech	SAL001	130
skeletal muscle	Clontech	SKM001	10 188
small intestine	Clontech	SIN001	10 49 59 70 86 94-96 102 119 159-160 164 183 189 268 272 300 303 307 318
spinal cord	Clontech	SPC001	4 29 43 65 79 83 106 149 158 172 176-177 189 211 265 324
stomach	Clontech	STO001	10 292
thalamus	Clontech	THA002	13 245 278 295 314
thymus	Clontech	THM001	4-5 149 258 302 325
thymus	Clontech	THMc02	6 13-14 25 96 125 151-152 211 221 243 258 270 303 306 314 317 325 330-331
thyroid gland	Clontech	THR001	25 65 75 108 149 172 182 189 193 236 270 291 303 305-306 312 320 324 326
trachea	Clontech	TRC001	1 142 297 303 317
umbilical cord	BioChain	FUC001	24 31 70 130 193 207 257 300 318 332
uterus	Clontech	UTR001	29 149 185 207 236
young liver	GIBCO	ALV001	182 314 331

*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human so\spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

Table 2A

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
337	gi12580867	Picea abies	60S ribosomal protein L13E	83	33
337	gi3127821	Drosophila subobscura	Sex-Peptide	66	41
337	gi3549864	Drosophila subobscura	Sex-peptide	66	41
338	AAY57951	Homo sapiens	Human transmembrane protein HTMPN-75.	77	33
338	gi642017	Hordeum vulgare	phospholipid transfer protein precursor	72	30
338	gi11037708	Triticum aestivum	lipid transfer protein precursor	72	34
339	AAY20852	Homo sapiens	Human neurofilament-H mutant protein fragment 11.	108	38
339	gi1888411	Homo sapiens	mRNA encoding chimaeric transcript of collagen type 1 alpha 1 and platelet derived growth factor beta, 314 bp.	80	30
339	AAW18664	Homo sapiens	Fragmented human NF-H gene +1 frameshift mutant product.	100	38
340	AAB08912	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:69.	251	100
340	gi12248917	Homo sapiens	mRNA for spinesin, complete cds.	251	100
340	AAB11699	Homo sapiens	Human serine protease BSSP2 (hBSSP2), SEQ ID NO:10.	251	100
341	gi13990776	Gallus gallus	immunoglobulin lambda chain	67	43
341	gi1086714	Caenorhabditis elegans	coded for by C. elegans cDNA yk74c8.5; Similar to small type-II membrane antigen	55	45
341	gi1469906	Gallus gallus	beta-1,4-galactosyltransferase	56	46
342	AAY17526	Homo sapiens	Human secreted protein clone AM349 2 protein.	1131	100
342	AAY02361	Homo sapiens	Polypeptide identified by the signal sequence trap method.	1131	100
342	AAW52834	Homo sapiens	Secreted protein encoded by clone AM349_2.	664	100
343	gi5579130	Hepatitis E virus	non-structural polyprotein	71	37
343	gi330005	Hepatitis E virus	poly-proline hinge	58	35
343	gi7768740	Homo sapiens	genomic DNA, chromosome 21q, section 89/105.	82	29
344	AAY86234	Homo sapiens	Human secreted protein HNTNC20, SEQ ID NO:149.	476	60
344	AAB24074	Homo sapiens	Human PRO1153 protein sequence SEQ ID NO:49.	111	46
344	AAY66735	Homo sapiens	Membrane-bound protein PRO1153.	111	46
345	gi12836893	Gallus gallus	IPR328-like protein	165	30
345	gi13357180	Homo sapiens	calcium channel gamma subunit 8 (CACNG8) mRNA, partial cds.	125	28
345	gi4558766	Homo sapiens	neuronal voltage gated calcium channel gamma-3 subunit mRNA, complete cds.	158	30
346	AAY79384	Homo sapiens	Human G protein coupled	396	100

Table 2A

120

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			receptor SLGP 7 transmembrane region.		
346	gi11225483	Homo sapiens	ETL protein (ETL) mRNA, complete cds.	396	100
346	AAB61144	Homo sapiens	Human NOV14 protein.	396	100
347	gi13195147	Mus musculus	HCH	209	77
347	gi1339910	Homo sapiens	Human DOCK180 protein mRNA, complete cds.	95	43
347	AAW03515	Homo sapiens	Human DOCK180 protein.	95	43
348	gi10176829	Arabidopsis thaliana	gene_id:MBB18.16~	79	32
349	gi10438431	Homo sapiens	cDNA: FLJ22155 fis, clone HRC00205.	518	34
349	gi10437336	Homo sapiens	cDNA: FLJ21267 fis, clone COL01717.	506	36
349	AAY07754	Homo sapiens	Human secreted protein fragment encoded from gene 11.	291	37
350	gi1552496	Homo sapiens	Human germline T-cell receptor beta chain Dopamine-beta-hydroxylase-like, TRY1, TRY2, TRY3, TCRBV27S1P, TCRBV22S1A2N1T, TCRBV9S1A1T, TCRBV7S1A1N2T, TCRBV5S1A1T, TCRBV13S3, TCRBV6S7P, TCRBV7S3A2T, TCRBV13S2A1T, TCRBV9S2A2PT, TCRBV7S2A1N4T, TCRBV13S9/13S2A1T, TCRBV6S5A1N1, TCRBV30S1P, TCRBV31S1, TCRBV13S5, TCRBV6S1A1N1, TCRBV32S1P, TCRBV5S5P, TCRBV1S1A1N1, TCRBV12S2A1T, TCRBV21S1, TCRBV8S4P, TCRBV12S3, TCRBV21S3A2N2T, TCRBV8S5P, TCRBV13S1 genes from bases 1 to 267156 (section 1 of 3).	614	100
350	gi33560	Homo sapiens	Human mRNA for T-cell receptor V beta gene segment V-beta-9, clone IGRb20.	609	100
350	gi37634	Homo sapiens	H.sapiens rearranged TCR Vbeta 9.1 mRNA for T cell receptor.	609	100
351	gi13960126	Homo sapiens	Similar to leucine-rich neuronal protein, clone MGC:4126, mRNA, complete cds.	162	80
351	gi14043281	Homo sapiens	clone IMAGE:3528313, mRNA, partial cds.	133	64
351	gi3135309	Homo sapiens	chromosome 7q22 sequence, complete sequence.	133	64
352	AAB61141	Homo sapiens	Human NOV11 protein.	370	86

Table 2A

121

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
352	gi4760778	Mus musculus	Ten-m2	369	100
352	gi5712201	Rattus norvegicus	neurestin alpha	369	100
353	AAW88628	Homo sapiens	Secreted protein encoded by gene 95 clone HPWAN23.	78	30
353	AAY57923	Homo sapiens	Human transmembrane protein HTMPN-47.	78	30
353	gi7109072	Plasmodium falciparum	PfEMP1 protein	78	37
354	gi1061424	Homo sapiens	Human PMS2 related (hPMSR3) gene, complete cds.	194	48
354	gi5738553	Homo sapiens	mRNA for zinc finger protein, clone cZNF41.5, partial.	175	48
354	gi5738547	Homo sapiens	mRNA for zinc finger protein, clone cZNF41.2, partial.	174	71
355	gi14161140	Streptococcus pyogenes	M protein	75	35
355	gi472917	Enterococcus hirae	v-type Na-ATPase	64	37
355	AAW00946	Homo sapiens	Human c-Fos protein.	63	40
356	gi6088092	Mesocricetus auratus	cytochrome P450	92	47
356	AAY91348	Homo sapiens	Human secreted protein sequence encoded by gene 3 SEQ ID NO:69.	130	40
356	gi4249595	Mus musculus	CYP2C40	115	34
357	gi12053357	Homo sapiens	mRNA; cDNA DKFZp586G2122 (from clone DKFZp586G2122); complete cds.	488	67
357	AAY27649	Homo sapiens	Human secreted protein encoded by gene No. 83.	62	35
357	gi9755390	Arabidopsis thaliana	F17F8.22	81	46
358	gi6273399	Homo sapiens	melanoma-associated antigen MG50 mRNA, partial cds.	359	95
358	AAW81030	Homo sapiens	Melanoma associated antigen MG50.	359	95
358	AAY70469	Homo sapiens	Human p53 target molecule, PRG2 protein.	359	95
359	gi7380324	Neisseria meningitidis Z2491	ClpB protein	91	32
359	gi7226713	Neisseria meningitidis MC58	clpB protein	91	32
359	gi9658311	Vibrio cholerae	integrase-related protein	61	34
360	AAB24074	Homo sapiens	Human PRO1153 protein sequence SEQ ID NO:49.	1023	99
360	AAY66735	Homo sapiens	Membrane-bound protein PRO1153.	1023	99
360	AAB65258	Homo sapiens	Human PRO1153 (UNQ583) protein sequence SEQ ID NO:351.	1023	99
361	gi1364247	Sus scrofa	Ca(2+)-transport ATPase (AA 989-1042); non-muscle isoform (1 is 3rd base in codon)	57	38
361	AAB65991	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 131.	73	34
361	AAB65992	Homo sapiens	Human secreted protein BLAST	73	34

Table 2A

122

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			search protein SEQ ID NO: 132.		
362	gi2150146	Mus musculus	sulfonylurea receptor 2A	634	73
362	gi8843832	Rattus norvegicus	sulphonylurea receptor 2b	375	73
362	gi3127175	Homo sapiens	sulfonylurea receptor 2A (SUR2) gene, alternatively spliced product, exon 38a and complete cds.	372	74
363	gi4467773	Helicobacter pylori	cytotoxin associated protein A	60	34
363	gi7248699	Helicobacter pylori	cytotoxin associated protein CagA	60	34
363	gi5851989	Helicobacter pylori	cytotoxin associated protein A	59	31
364	gi13278675	Homo sapiens	clone MGC:11170, mRNA, complete cds.	77	41
364	gi6457690	Deinococcus radiodurans	2-oxo acid dehydrogenase, E2 component	90	31
364	gi179521	Homo sapiens	Human bullous pemphigoid (BP180) mRNA, partial cds.	72	36
365	AAB52176	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 132.	468	95
365	AAR27651	Homo sapiens	Human calcium channel 27980/13.	117	26
365	gi179764	Homo sapiens	Human neuronal DHP-sensitive, voltage-dependent, calcium channel alpha-1D subunit mRNA, complete cds.	117	26
366	gi13623421	Homo sapiens	Similar to RIKEN cDNA 5730589L02 gene, clone MGC:13124, mRNA, complete cds.	495	98
366	gi12803383	Homo sapiens	clone MGC:2099, mRNA, complete cds.	189	100
366	gi13111983	Homo sapiens	clone MGC:4221, mRNA, complete cds.	189	100
367	AAW75100	Homo sapiens	Human secreted protein encoded by gene 44 clone HE8CJ26.	121	83
367	gi11275978	Homo sapiens	NOTCH 2 (N2) mRNA, complete cds.	125	87
367	AAY06816	Homo sapiens	Human Notch2 (humN2) protein sequence.	125	87
368	gi2696709	Mus musculus	RST	258	43
368	gi2687858	Pseudopleuronectes americanus	renal organic anion transporter	236	40
368	gi4586315	Homo sapiens	ORCTL3 mRNA for organic-cation transporter like 3, complete cds.	232	37
369	gi11463949	Homo sapiens	hUGTrel7 mRNA for UDP-glucuronic acid, complete cds.	256	100
369	AAB60119	Homo sapiens	Human transport protein TPPT-39.	175	63
369	AAB56473	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1051.	175	63
370	gi3986168	Lentinula edodes	SHP1	55	31
370	gi12805659	Mus musculus	Similar to syndecan 4	53	34

Table 2A

123

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
371	AAB88377	Homo sapiens	Human membrane or secretory protein clone PSEC0113.	370	94
371	gi12656637	Mus musculus	equilibrative nucleoside transporter 3	109	25
371	gi3877156	Caenorhabditis elegans	F44D12.9	92	32
372	gi9828006	Leishmania major	probable ctg26 alteRNAte open reading frame	60	40
372	gi4096496	Homo sapiens	Human pre-B cell Ig heavy chain mRNA, third complementarity-determining region, clone PBT-55, partial cds.	55	47
372	gi3005708	Homo sapiens	clone 23619 phosphoprotein mRNA, partial cds.	66	33
373	gi1339910	Homo sapiens	Human DOCK180 protein mRNA, complete cds.	121	54
373	AAW03515	Homo sapiens	Human DOCK180 protein.	121	54
373	gi13195147	Mus musculus	HCH	107	61
374	gi11036344	Pichia canadensis	NADH dehydrogenase subunit 4L	69	38
374	gi10175432	Bacillus halodurans	D-alanine aminotransferase	87	35
374	gi10639223	Thermoplasma acidophilum	ethanolamine permease related protein	88	27
375	AAB90654	Homo sapiens	Human secreted protein, SEQ ID NO: 197.	58	29
375	AAY36085	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 470.	56	34
375	gi3617829	Gallus gallus	gallinacin 1 prepropeptide	55	42
376	gi14189735	Homo sapiens	ATP-binding cassette transporter family A member 12 (ABCA12) mRNA, complete cds.	251	43
376	gi14209834	Mus musculus	ATP-binding cassette transporter sub-family A member 7	199	39
376	gi9211112	Homo sapiens	macrophage ABC transporter (ABCA7) mRNA, complete cds.	196	40
377	gi8919747	Cottontail rabbit papillomavirus	e8	65	36
377	gi8919568	Cottontail rabbit papillomavirus	E8	64	36
377	gi5679184	Xanthomonas campestris pv. glycines	HrcU homolog	80	25
378	AAY30817	Homo sapiens	Human secreted protein encoded from gene 7.	569	98
378	gi3411233	Mus musculus	IER5	107	37
378	AAG02396	Homo sapiens	Human secreted protein, SEQ ID NO: 6477.	85	61
379	AAY99353	Homo sapiens	Human PRO1415 (UNQ731) amino acid sequence SEQ ID NO:50.	1435	99
379	AAB88426	Homo sapiens	Human membrane or secretory protein clone PSEC0199.	1428	99

Table 2A

124

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
379	gi11230635	Homo sapiens	CD30 gene for cytokine receptor CD30, exons 1-8.	106	29
380	gi6636340	Rattus norvegicus	myosin heavy chain Myr 8	157	61
380	gi10863773	Rattus norvegicus	myosin heavy chain Myr 8b	157	61
380	AAB51865	Homo sapiens	Human secreted protein sequence encoded by gene 39 SEQ ID NO:98.	71	31
381	gi9789476	Mus musculus	claudin-19	98	41
381	gi3335182	Mus musculus	claudin-1	98	32
381	gi12805093	Mus musculus	claudin 1	98	32
382	gi213109	Discopyge ommata	synaptic vesicle protein	75	36
382	gi1679584	Cavia porcellus	membrane cofactor protein precursor	80	37
382	gi1655471	Cavia porcellus	membrane cofactor protein(GMP1-full)	80	37
383	gi14330016	Mus musculus	bM401L17.2.1 (cholinergic receptor, nicotinic, alpha polypeptide 4 (isoform 1))	164	50
383	gi9886085	Mus musculus	nicotinic acetlycholine receptor alpha 4 subunit	164	50
383	gi14330017	Mus musculus	bM401L17.2.2 (cholinergic receptor, nicotinic, alpha polypeptide 4 (isoform 2))	164	50
384	gi409995	Rattus sp.	mucin	137	47
384	gi4995986	Human herpesvirus 6	13.6% identical to DR8 gene of strain U1102 of HHV-6	135	32
384	gi2388546	Homo sapiens	Human Xq28 BAC RP11-159I8 (Roswell Park Cancer Institute Human BAC Library), Cosmid LL0XNC01-3C3 (LLNL X Chromsome Library), and BAC GS1-92B2 (Genome Systems Human BAC Library) complete sequence.	118	37
385	AAY58174	Homo sapiens	Human embryogenesis protein, EMPRO.	872	96
385	gi3879940	Caenorhabditis elegans	Similarity to Mouse H(beta)58 protein (SW:HB58_MOUSE)	650	67
385	gi3342000	Homo sapiens	H beta 58 homolog	666	70
386	gi13359817	Escherichia coli O157:H7	high-affinity choline transport	1021	100
386	gi1657512	Escherichia coli	high-affinity choline transport protein	1021	100
386	gi1786506	Escherichia coli K12	high-affinity choline transport	1021	100
387	gi10584129	Halobacterium sp. NRC-1	Vng6071c	81	27
387	gi10584473	Halobacterium sp. NRC-1	Vng6455c	81	27
387	gi12723038	Lactococcus lactis subsp. lactis	UNKNOWN PROTEIN	58	28
388	gi13364609	Escherichia coli O157:H7	fumarate reductase FrdD	515	96
388	gi145266	Escherichia coli	g13 protein	515	96
388	gi1790594	Escherichia coli K12	fumarate reductase, anaerobic,	515	96

Table 2A

125

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			membrane anchor polypeptide		
389	gi1160319	Escherichia coli	aldohexuronate transport system	928	96
389	gi13363448	Escherichia coli O157:H7	transport protein of hexuronates	928	96
389	gi2367193	Escherichia coli K12	transport of hexuronates	928	96
390	gi395270	Escherichia coli	FepE	402	100
390	gi1786802	Escherichia coli K12	ferric enterobactin (enterochelin) transport	402	100
390	gi1778503	Escherichia coli	ferric enterobactin transport protein	402	100
391	gi145521	Escherichia coli	methyl-accepting chemotaxis protein II	411	73
391	gi1736539	Escherichia coli	Methyl-accepting chemotaxis protein II (MCP-II) (Aspartate chemoreceptor protein).	411	73
391	gi1788195	Escherichia coli K12	methyl-accepting chemotaxis protein II, aspartate sensor receptor	411	73
392	AAB37990	Homo sapiens	Human secreted protein encoded by gene 7 clone HWLHH15.	303	98
392	gi312188	Bovine herpesvirus 1	glycoprotein gD	85	29
392	gi5668989	Bovine herpesvirus type 1.1	glycoprotein D precursor	76	29
393	gi14456429	Equus caballus	galanin receptor 1	69	28
393	gi3282259	Cucumaria pseudocurata	ND4L	69	30
393	gi3282257	Cucumaria miniata	ND4L	68	30
394	gi3702702	bacteriophage Vf33	Vpf77	65	30
394	gi3702711	bacteriophage Vf12	Vpf77	65	30
394	gi1742947	Alcaligenes sp.	urf-1 (merE)	64	31
395	gi263516	Azospirillum brasilense, Sp7, Peptide Partial, 70 aa	NifB {N-terminal}	58	39
395	gi9622741	Conus catus	four-loop conotoxin precursor	57	33
395	gi149569	Lactobacillus sp.	lactacin F	56	40
396	gi896286	Leishmania tarentolae	NH2 terminus uncertain	123	19
396	gi4155384	Helicobacter pylori J99	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN	120	27
396	gi1542807	Asterina pectinifera	NADH-dehydrogenase subunit 4L	98	27
397	AAB88433	Homo sapiens	Human membrane or secretory protein clone PSEC0210.	299	55
397	gi6996444	Homo sapiens	CTL2 gene.	299	55
397	AAB24284	Homo sapiens	Human H38087 (clone GTB6) protein sequence SEQ ID NO:7.	295	54
398	gi6807868	Homo sapiens	mRNA; cDNA DKFZp434G0625 (from clone DKFZp434G0625); partial cds.	324	68
398	AAY13373	Homo sapiens	Amino acid sequence of protein PRO235.	209	62
398	AAB33420	Homo sapiens	Human PRO235 protein UNQ209 SEQ ID NO:31.	209	62

Table 2A

126

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
399	gi10434911	Homo sapiens	cDNA FLJ13068 fis, clone NT2RP3001739, weakly similar to HYPOTHETICAL 72.5 KD PROTEIN C2F7.10 IN CHROMOSOME I.	573	100
399	gi7022673	Homo sapiens	cDNA FLJ10562 fis, clone NT2RP2002701.	109	43
399	AAY87090	Homo sapiens	Human secreted protein sequence SEQ ID NO:129.	109	43
400	AAB63630	Homo sapiens	Human gastric cancer associated antigen protein sequence SEQ ID NO:992.	165	55
400	AAB63629	Homo sapiens	Human gastric cancer associated antigen protein sequence SEQ ID NO:991.	170	55
400	AAR06471	Homo sapiens	Derived protein from clone ICA525 (ATCC 40704).	172	55
401	gi13543949	Homo sapiens	Similar to RIKEN cDNA 2810432L12 gene, clone MGC:12992, mRNA, complete cds.	2104	100
401	AAY87340	Homo sapiens	Human signal peptide containing protein HSPP-117 SEQ ID NO:117.	2104	100
401	gi3876730	Caenorhabditis elegans	F35C11.4	181	27
402	gi5001993	Dissostichus mawsoni	chimeric AFGP/trypsinogen-like serine protease precursor	199	49
402	gi295736	Dictyostelium discoideum	spore coat protein sp96	189	48
402	gi2114321	Equine herpesvirus 1	membrane glycoprotein	186	39
403	gi7239364	Homo sapiens	acetylcholinesterase collagen-like tail subunit (COLQ) gene, exon 17; and complete cds, alternatively spliced.	136	29
403	gi3599478	Acanthamoeba castellanii	Myosin-IA	137	35
403	gi3858883	Acanthamoeba castellanii	myosin I heavy chain kinase	133	30
404	AAB66272	Homo sapiens	Human TANGO 378 SEQ ID NO: 29.	664	89
404	gi6006811	Mus musculus	serpentine receptor	261	40
404	AAB01247	Homo sapiens	Human HE6 receptor.	263	38
405	gi13623515	Homo sapiens	clone MGC:12705, mRNA, complete cds.	94	87
405	gi1017781	bacteriophage lambda	Rz1 protein precursor	44	41
405	gi6599136	Homo sapiens	mRNA; cDNA DKFZp434F216 (from clone DKFZp434F216); partial cds.	94	87
406	AAC84384_aa1	Homo sapiens	Human A236 polypeptide coding sequence.	693	100
406	gi10438797	Homo sapiens	cDNA: FLJ22415 fis, clone HRC08561.	692	100
406	AAY41692	Homo sapiens	Human PRO 363 protein sequence.	692	100

Table 2A

127

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
407	gi8515813	Rattus norvegicus	RSD-6	84	25
407	gi12657809	Simian immunodeficiency virus	gag protein	83	25
407	gi9454456	Human immunodeficiency virus type 1	pol protein	60	35
408	AAY71056	Homo sapiens	Human membrane transport protein, MTRP-1.	143	76
408	gi13096889	Mus musculus	Similar to ATPas, class II, type 9B	142	68
408	gi13905302	Mus musculus	Similar to ATPase, class II, type 9A	119	63
409	gi2384752	Paracentrotus lividus	transcription factor; PaxA	56	47
409	gi6601486	Ovis aries	pulmonary surfactant protein B	76	30
409	AAR41266	Homo sapiens	vWF fragment Arg441-Tyr508, deltaCys474-Pro488.	56	47
410	AAY99420	Homo sapiens	Human PRO1486 (UNQ755) amino acid sequence SEQ ID NO:287.	1082	100
410	AAW88747	Homo sapiens	Secreted protein encoded by gene 45 clone HCESF40.	1069	99
410	gi6942096	Mus musculus	CBLN3	942	94
411	gi11558496	Sus scrofa	sodium iodide symporter	170	51
411	gi12642414	Mus musculus	sodium iodide symporter NIS	184	39
411	gi14290145	Mus musculus	sodium iodide symporter	184	39
412	AAY66645	Homo sapiens	Membrane-bound protein PRO1310.	554	100
412	AAB65168	Homo sapiens	Human PRO1310 protein sequence SEQ ID NO:62.	554	100
412	gi2921092	Mus musculus	carboxypeptidase X2	281	58
414	gi5901822	Drosophila melanogaster	EG:118B3.2	160	70
414	AAB29877	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 135.	127	52
414	AAB29878	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 136.	121	41
415	gi58442	Human adenovirus type 41	8.0K protein (AA 1-74)	56	44
415	gi388253	Trifolium repens	ribulose bisphosphate carboxylase	54	32
415	gi1345574	Sinapis alba	small subunit ribulose 1,5-bisphosphate carboxylase (AA 1-82)	57	36
416	gi3047402	Homo sapiens	monocarboxylate transporter 2 (hMCT2) mRNA, complete cds.	539	34
416	gi7688756	Mus musculus	monocarboxylate transporter 4	296	48
416	gi3834395	Homo sapiens	monocarboxylate transporter 2 (MCT2) mRNA, complete cds.	528	33
417	gi6136782	Mus musculus	synaptotagmin V	595	91
417	gi14210264	Rattus norvegicus	synaptotagmin 5	592	91
417	gi6136792	Mus musculus	synaptotagmin X	268	43
418	AAB53400	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID	493	100

Table 2A

128

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO:940.		
418	gi6760350	Homo sapiens	cytomegalovirus partial fusion receptor mRNA, partial cds.	348	98
418	gi603380	Saccharomyces cerevisiae	Yer140wp	106	30
419	AAB12136	Homo sapiens	Hydrophobic domain protein from clone HP10625 isolated from Liver cells.	1142	100
419	AAB24036	Homo sapiens	Human PRO4407 protein sequence SEQ ID NO:47.	1142	100
419	AAY57952	Homo sapiens	Human transmembrane protein HTMPN-76.	1142	100
420	gi2654984	Hepatitis GB virus C	polyprotein	50	38
420	gi861305	Caenorhabditis elegans	similar to C. elegans protein F59B2.2	75	32
420	AAW75055	Homo sapiens	Fragment of human secreted protein encoded by gene 18.	52	38
421	gi2696709	Mus musculus	RST	95	47
421	gi1293672	Mus musculus	kidney-specific transport protein	93	40
421	gi7707622	Homo sapiens	hOAT4 mRNA for organic anion transporter 4, complete cds.	93	37
422	gi17829	Brassica napus	LEA76 peptide (AA 1-280)	137	27
422	gi11994339	Arabidopsis thaliana	embryonic abundant protein LEA-like	119	28
422	gi3873646	Caenorhabditis elegans	AC3.3	123	27
423	AAB74753	Homo sapiens	Human secreted protein sequence encoded by gene 21 SEQ ID NO:62.	38	54
423	gi2369777	Drosophila mauritiana	sex-peptide	39	53
423	gi2369804	Drosophila simulans	sex-peptide	39	53
424	gi13959739	Caprine arthritis-encephalitis virus	envelope glycoprotein	87	33
424	gi5732606	Hepatitis B virus	precore/core mutant protein	74	33
424	gi4033542	Hepatitis B virus	truncated pre-core-protein	72	34
425	AAB53400	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:940.	220	91
425	gi1177469	Homo sapiens	gene for interleukin-10.	37	46
425	AAB62192	Homo sapiens	Human interleukin-10 (IL-10) protein.	37	46
426	gi1336041	Homo sapiens	Human olfactory receptor (OLF1) gene, complete cds.	482	50
426	gi1246530	Gallus gallus	olfactory receptor 2	474	50
426	gi1246534	Gallus gallus	olfactory receptor 4	474	50
427	AAY36243	Homo sapiens	Human secreted protein encoded by gene 20.	64	48
427	gi409995	Rattus sp.	mucin	65	57
427	gi11141770	Bos taurus	Toll-like receptor 4	80	29
428	gi8918871	Plasmid F	96 pct identical to gp:AB021078_30	288	98
428	gi4512467	Plasmid CoIIb-P9	100 pct identical to 25 residues	256	93

Table 2A

129

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			of 79 aa protein sp:YPF8_ECOLI		
428	gi47517	Synechocystis sp. PCC 6803	ATPase subunit epsilon	72	45
429	gi5139695	Cucumis sativus	expressed in cucumber hypocotyls	85	28
429	gi3406819	Mus musculus	growth factor receptor	63	47
429	AAG03497	Homo sapiens	Human secreted protein, SEQ ID NO: 7578.	61	51
430	AAB18985	Homo sapiens	Amino acid sequence of a human transmembrane protein.	251	35
430	gi6013381	Rattus norvegicus	TM6P1	246	33
430	AAE00330	Homo sapiens	Human membrane-bound protein-60 (Zsig60).	251	35
432	gi1046315	Plasmodium vivax	merozoite surface protein-1	88	34
432	gi2213834	Plasmodium vivax	merozite surface protein 1	85	29
432	gi537916	Lilium longiflorum	meiotin-1	87	32
433	AAY91618	Homo sapiens	Human secreted protein sequence encoded by gene 20 SEQ ID NO:291.	63	29
433	AAG02988	Homo sapiens	Human secreted protein, SEQ ID NO: 7069.	58	29
434	gi220411	Mus musculus	N-methyl-D-aspartate receptor channel subunit epsilon 1	159	100
434	gi286234	Rattus norvegicus	N-methyl-D-aspartate receptor subunit	159	100
434	gi2155310	Rattus norvegicus	N-methyl-D-aspartate receptor NMDAR2A subunit; NMDA receptor NMDAR2A subunit	159	100
435	AAB66267	Homo sapiens	Human TANGO 272 SEQ ID NO: 14.	697	50
435	AAY72712	Homo sapiens	HTLIH44 clone human attractin-like protein.	570	47
435	AAY72715	Homo sapiens	HFICU08 clone human attractin-like protein.	565	47
436	gi2589210	Mus musculus	calcium-sensing receptor related protein 3	105	35
436	gi3130157	Takifugu rubripes	pheromone receptor	106	34
436	gi2589208	Mus musculus	calcium-sensing receptor related protein 2	99	33
437	gi2384746	Mus musculus	testicular condensing enzyme	681	52
437	gi4633135	Mus musculus	condensing enzyme	681	52
437	gi12652723	Homo sapiens	clone MGC:3295, mRNA, complete cds.	276	29
438	gi12224992	Homo sapiens	mRNA; cDNA DKFZp667O2416 (from clone DKFZp667O2416).	877	100
438	gi4929647	Homo sapiens	CGI-89 protein mRNA, complete cds.	603	61
438	gi12652585	Homo sapiens	CGI-89 protein, clone MGC:845, mRNA, complete cds.	602	60
439	AAY36047	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO.	61	57

Table 2A

130

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			432.		
439	AAG01318	Homo sapiens	Human secreted protein, SEQ ID NO: 5399.	59	44
439	AAW74979	Homo sapiens	Human secreted protein encoded by gene 105 clone HSVAF07.	58	35
440	gi12314108	Homo sapiens	Human DNA sequence from clone RP1-230I3 on chromosome 6q22.1-22.33 Contains part of a gene for a novel protein, STSs and GSSs, complete sequence.	634	100
440	gi10434835	Homo sapiens	cDNA FLJ13018 fis, clone NT2RP3000685.	435	68
440	gi1491712	Homo sapiens	H.sapiens mRNA for novel protein.	95	56
442	gi861305	Caenorhabditis elegans	similar to C. elegans protein F59B2.2	124	30
442	gi10177114	Arabidopsis thaliana	amino acid transporter protein-like	91	34
442	gi2576363	Arabidopsis thaliana	amino acid transport protein	79	29
443	AAY28678	Homo sapiens	Human cw272_7 secreted protein.	324	38
443	gi13185723	Homo sapiens	n 1755 can be A, G, C, or T	248	30
443	AAB70537	Homo sapiens	Human PRO7 protein sequence SEQ ID NO:14.	248	30
444	gi10186503	Homo sapiens	sialic acid-specific acetylesterase II mRNA, complete cds, alternatively spliced.	932	100
444	gi6808138	Homo sapiens	mRNA; cDNA DKFZp761A051 (from clone DKFZp761A051); partial cds.	923	100
444	gi10242345	Homo sapiens	sialic acid-specific 9-O-acetylesterase I mRNA, complete cds.	753	100
445	gi7328084	Homo sapiens	mRNA; cDNA DKFZp761L0812 (from clone DKFZp761L0812); partial cds.	225	82
445	gi7576817	Plasmodium falciparum	merozoite surface protein 2	94	38
445	gi3261822	Mycobacterium tuberculosis	PE_PGRS	103	36
446	gi3165565	Caenorhabditis elegans	contains similarity to transmembrane domains found in HMG CoA reductases and drosophila patched protein (SW:P18502)	129	25
446	gi1825729	Caenorhabditis elegans	similar to drosophila membrane protein PATCHED SP:P18502 (PID:g129645)	125	26
446	gi15120	enterobacteria phage P1	unidentified reading frame	67	31
447	AAB88481	Homo sapiens	Human membrane or secretory protein clone PSEC0251.	254	73
447	gi57115	Rattus norvegicus	ribosomal protein L31 (AA 1-125)	175	67

Table 2A

131

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
447	gi14198321	Mus musculus	ribosomal protein L31	175	67
448	gi3130189	Takifugu rubripes	pheromone receptor	212	63
448	gi2589208	Mus musculus	calcium-sensing receptor related protein 2	205	50
448	gi2589210	Mus musculus	calcium-sensing receptor related protein 3	203	48
449	gi13452508	Mus musculus	claudin 14	438	40
449	gi12597447	Homo sapiens	claudin 14 (CLDN14) mRNA, complete cds.	438	39
449	gi7768724	Homo sapiens	genomic DNA, chromosome 21q, section 70/105.	438	39
450	AAR12603	Homo sapiens	SIB 121 intestinal mucin.	148	53
450	AAW36946	Homo sapiens	Protein encoded by 5' fragment of clone M8_2.	92	35
450	AAY91378	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:99.	86	45
451	gi13561518	Homo sapiens	GalNAc-4-sulfotransferase 2 mRNA, complete cds, alternatively spliced.	213	97
451	gi12711481	Homo sapiens	N-acetylgalactosamine 4-O-sulfotransferase 2 GalNAc4ST-2 mRNA, complete cds.	187	97
451	AAY86315	Homo sapiens	Human secreted protein HNTMX29, SEQ ID NO:230.	63	27
452	gi3150438	Human endogenous retrovirus K	pol-env	264	51
452	gi3150441	Human endogenous retrovirus K	envelope protein	258	50
452	gi5802817	Homo sapiens	endogenous retrovirus HERV-K104 long terminal repeat, complete sequence; and Gag protein (gag) and envelope protein (env) genes, complete cds.	258	51
453	AAY91625	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	547	97
453	AAU00437	Homo sapiens	Human dendritic cell membrane protein FIRE.	547	97
453	AAW30638	Homo sapiens	Partial human 7-transmembrane receptor HAPO167 protein.	374	66
454	AAY96963	Homo sapiens	Wound healing tissue peptidoglycan recognition protein-like protein.	1811	92
454	AAY96962	Homo sapiens	Keratinocyte peptidoglycan recognition protein-like protein.	768	62
454	AAY76124	Homo sapiens	Human secreted protein encoded by gene 1.	768	62
455	AAB72286	Homo sapiens	Human ADAMTS-9 amino acid sequence.	1009	100
455	AAB72301	Homo sapiens	Human ADAMTS-9 alternative amino acid sequence.	1009	100
455	AAB90617	Homo sapiens	Human secreted protein, SEQ ID NO: 155.	358	39

Table 2A

132

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
456	gi4323581	Homo sapiens	senescence-associated epithelial membrane protein (SEMP1) mRNA, complete cds.	150	100
456	gi4559278	Homo sapiens	claudin-1 (CLDN1) mRNA, complete cds.	150	100
456	gi13383364	Homo sapiens	claudin-1 (CLDN1) gene, exon 4 and complete cds.	150	100
457	AAW93960	Homo sapiens	Human 53BP2:IP-2 protein fragment.	59	45
457	AAY19607	Homo sapiens	SEQ ID NO 325 from WO9922243.	57	64
457	AAY07942	Homo sapiens	Human secreted protein fragment encoded from gene 91.	55	42
458	gi4406172	Human herpesvirus 4	latent membrane protein-1	159	37
458	gi475574	Human herpesvirus 4 type 2	latent membrane protein 1	153	39
458	gi2736358	Caenorhabditis elegans	Contains similarity to Pfam domain: PF00069 (pkinase), Score=214.7, E-value=4.3e-61, N=1	155	51
459	AAB43892	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1337.	253	83
459	gi6456100	Mus musculus	F-box protein FBL10	247	83
459	gi14250563	Homo sapiens	clone IMAGE:3163445, mRNA, partial cds.	253	83
460	gi552087	Drosophila melanogaster	crumbs protein	127	45
460	AAY66747	Homo sapiens	Membrane-bound protein PRO1158.	67	46
460	AAB87559	Homo sapiens	Human PRO1158.	67	46
461	AAB39181	Homo sapiens	Human secreted protein sequence encoded by gene 3 SEQ ID NO:61.	57	41
462	AAW71565	Homo sapiens	Hepatocyte nuclear factor 4 alpha polypeptide (exon 2 product).	44	36
462	gi2804240	Rattus norvegicus	histidase	56	42
462	gi149163	Plasmid pJHC-MW1	streptomycin-spectinomycin resistance protein	65	71
463	gi10435833	Homo sapiens	cDNA FLJ13729 fis, clone PLACE3000121, weakly similar to VESICULAR TRAFFIC CONTROL PROTEIN SEC15.	233	100
463	gi6807998	Homo sapiens	mRNA; cDNA DKFZp76112124 (from clone DKFZp76112124); partial cds.	195	80
463	gi7023795	Homo sapiens	cDNA FLJ11251 fis, clone PLACE1008813.	195	80
464	gi5668598	Homo sapiens	Wiskott-Aldrich syndrome protein interacting protein (WASPIP) mRNA, partial cds.	156	33
464	gi1314755	Mus musculus	Wiskott-Aldrich Syndrome Protein	140	33
464	gi4096355	Mus musculus	Wiskott-Aldrich syndrome protein (WASP)	140	33

Table 2A

133

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
465	gi4886381	Human papillomavirus type 16	E5 protein	54	36
465	AAB28331	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 115.	54	36
465	gi4886413	Human papillomavirus type 16	E5 protein	53	26
466	gi12276062	Homo sapiens	group XII secreted phospholipase A2 mRNA, complete cds.	354	100
466	gi12276193	Homo sapiens	FKSG38 (FKSG38) mRNA, complete cds.	354	100
466	AAY88271	Homo sapiens	Human TANGO 180 protein.	354	100
467	gi4885010	Conus textile	O-superfamily conotoxin TxO5 precursor	73	26
467	gi6409400	Conus textile	conotoxin scaffold VI/VII precursor	71	25
467	AAW78192	Homo sapiens	Human secreted protein encoded by gene 67 clone HTOFC34.	67	39
468	AAB38330	Homo sapiens	Human secreted protein encoded by gene 10 clone HTEBV72.	214	97
468	gi2335059	Mus musculus	IgG receptor	76	52
468	gi969034	Mus musculus	Fc gamma receptor IIb1	76	52
469	gi13311009	Homo sapiens	NYD-SP16 mRNA, complete cds.	488	100
469	gi3287162	Human immunodeficiency virus type 1	vpu	69	26
469	gi1303982	Bacillus subtilis	YqkE	59	40
470	AAB13343	Homo sapiens	Human cortexin-like protein.	204	53
470	AAB38538	Homo sapiens	Human secreted protein sequence encoded by gene 17 SEQ ID NO:75.	57	39
470	AAB34316	Homo sapiens	Human secreted protein sequence encoded by gene 18 SEQ ID NO:77.	54	34
471	gi13938651	Mus musculus	Similar to conserved membrane protein at 44E	502	83
471	gi14194169	Arabidopsis thaliana	At1g05960/T21E18_20	124	30
471	gi265786	human, mRNA, 1271 nt	betacellulin . [Homo	75	57
472	gi310100	Rattus norvegicus	developmentally regulated protein	539	80
472	AAW52812	Homo sapiens	Human induced tumour protein.	227	37
472	AAY07771	Homo sapiens	Human secreted protein fragment encoded from gene 28.	221	40
473	AAY71294	Homo sapiens	Human orphan G protein-coupled receptor hRUP3.	1711	100
473	AAB02828	Homo sapiens	Human G protein coupled receptor hRUP3 protein SEQ ID NO:8.	1711	100
473	gi1204095	Takifugu rubripes	dopamine receptor	237	28
474	gi3041879	Mus musculus	LNXp80	556	54

Table 2A

134

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
474	gi3041881	Mus musculus	LNXp70	556	54
474	gi13183073	Homo sapiens	multi-PDZ-domain-containing protein mRNA, complete cds.	539	56
475	AAB08872	Homo sapiens	Amino acid sequence of a human secretory protein.	77	93
475	gi5734537	Methanothermobacter thermautotrophicus	transmembrane protein 9.0 kDa	62	43
475	gi13357178	Homo sapiens	calcium channel gamma subunit 7 (CACNG7) mRNA, complete cds.	78	38
476	gi5070458	tomato yellow leaf curl virus	BV2 protein	60	33
476	gi9944667	Amsacta moorei entomopoxvirus	AMV144	60	26
476	gi293853	Mus musculus	betacellulin	48	25
477	gi10799398	Homo sapiens	chromosome 19, BAC BC349142 (CTC-518B2), complete sequence.	1513	100
477	gi6063386	Homo sapiens	kallikrein-like protein 4 KLK-L4 gene, complete cds.	1513	100
477	gi4884462	Homo sapiens	mRNA; cDNA DKFZp586J1923 (from clone DKFZp586J1923); partial cds.	912	98
478	AAB90602	Homo sapiens	Human secreted protein, SEQ ID NO: 140.	704	100
478	AAB90662	Homo sapiens	Human secreted protein, SEQ ID NO: 205.	704	100
478	AAB90571	Homo sapiens	Human secreted protein, SEQ ID NO: 109.	700	99
479	AAB53436	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:976.	82	33
479	AAG02279	Homo sapiens	Human secreted protein, SEQ ID NO: 6360.	82	61
479	gi3879077	Caenorhabditis elegans	R10E11.9	81	35
480	gi581191	Escherichia coli	unidentified reading frame (AA 1-79)	64	36
480	gi929915	synthetic construct	insulin C chain	61	58
480	AAP60248	Homo sapiens	Human proinsulin.	61	58
481	AAB24074	Homo sapiens	Human PRO1153 protein sequence SEQ ID NO:49.	136	42
481	AAY66735	Homo sapiens	Membrane-bound protein PRO1153.	136	42
481	AAB65258	Homo sapiens	Human PRO1153 (UNQ583) protein sequence SEQ ID NO:351.	136	42
482	AAB08854	Homo sapiens	Amino acid sequence of a human secretory protein.	787	100
482	AAY87268	Homo sapiens	Human signal peptide containing protein HSPP-45 SEQ ID NO:45.	787	100
482	AAY66723	Homo sapiens	Membrane-bound protein PRO1100.	787	100

Table 2A

135

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
483	gi14211714	Homo sapiens	naked cuticle-1 (NKD1) mRNA, complete cds.	193	92
483	AAB08216	Homo sapiens	A protein related to Drosophila naked cuticle polypeptide.	193	92
483	gi13487305	Mus musculus	Nkd	151	62
484	gi3452275	Pleuronectes americanus	aminopeptidase N	215	28
484	gi2766187	Gallus gallus	aminopeptidase Ey	178	32
484	gi3776238	Rattus norvegicus	aminopeptidase N	151	29
485	AAB58305	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 643.	273	100
485	gi5830684	variola minor virus	A20L protein	57	24
485	gi297302	Variola virus	A19L	57	24
486	AAB38019	Homo sapiens	Human secreted protein encoded by gene 27 clone HPJBF63.	583	99
486	AAB38010	Homo sapiens	Human secreted protein encoded by gene 27 clone HOUHD63.	576	98
486	gi167020	Hordeum vulgare	C-hordein storage protein	47	27
487	AAY91385	Homo sapiens	Human secreted protein sequence encoded by gene 40 SEQ ID NO:106.	969	100
487	gi4126441	Homo sapiens	CD22 gene variant 6, partial cds.	68	34
487	gi201798	Mus musculus	T-cell receptor beta	95	29
488	gi9971734	Galleria mellonella	heavy-chain fibroin	121	34
488	gi3002791	Homo sapiens	macrophage receptor MARCO mRNA, complete cds.	81	28
488	gi5231092	Homo sapiens	macrophage receptor (MARCO) gene, exon 17 and complete cds.	81	28
489	gi409995	Rattus sp.	mucin	173	64
489	gi4063042	Cryptosporidium parvum	GP900; mucin-like glycoprotein	134	38
489	gi5732924	Toxocara canis	excretory/secretory mucin MUC-4	112	29
490	gi1841555	Homo sapiens	HLA class III region containing NOTCH4 gene, partial sequence, homeobox PBX2 (HPBX) gene, receptor for advanced glycosylation end products (RAGE) gene, complete cds, and 6 unidentified cds, complete sequence.	422	100
490	AAB25697	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:86.	122	40
490	AAB25755	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:144.	122	40
491	gi5732924	Toxocara canis	excretory/secretory mucin MUC-4	114	34
491	gi5732920	Toxocara canis	excretory/secretory mucin MUC-2	113	32
491	gi409995	Rattus sp.	mucin	95	29
492	AAB70534	Homo sapiens	Human PRO4 protein sequence	395	100

Table 2A

136

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SEQ ID NO:8.		
492	AAY13377	Homo sapiens	Amino acid sequence of protein PRO257.	395	100
492	AAB80245	Homo sapiens	Human PRO257 protein.	395	100
493	gi12656447	Plasmodium falciparum	erythrocyte membrane protein 1	73	33
493	AAG04067	Homo sapiens	Human secreted protein, SEQ ID NO: 8148.	73	51
493	gi4200249	Homo sapiens	H.sapiens gene from PAC 747L4.	76	32
494	gi12003279	Perilla frutescens	15kD oleosin-like protein 1	77	36
494	gi409424	Homo sapiens	Human carboxyl ester lipase like protein (CELL) mRNA, complete cds.	59	32
494	gi609286	Xenopus laevis	xsna	79	30
495	gi1841555	Homo sapiens	HLA class III region containing NOTCH4 gene, partial sequence, homeobox PBX2 (HPBX) gene, receptor for advanced glycosylation end products (RAGE) gene, complete cds, and 6 unidentified cds, complete sequence.	80	42
495	AAB18976	Homo sapiens	Amino acid sequence of a human transmembrane protein.	69	40
495	AAW73192	Homo sapiens	Human vesicle trafficking protein.	43	38
496	gi13241972	Mus musculus	SugarCrisp	841	56
496	gi13241970	Gallus gallus	SugarCrisp	840	59
496	gi2943716	Homo sapiens	mRNA for 25 kDa trypsin inhibitor, complete cds.	840	63
497	gi4584539	Arabidopsis thaliana	extensin-like protein	138	34
497	gi306316	Herpesvirus papio	EBNA-2	171	38
497	gi1632787	Human herpesvirus 4	BYRF1, encodes EBNA-2 (Dambaugh et al, 1984; Dillner et al, 1984)	142	35
498	gi13185723	Homo sapiens	n 1755 can be A, G, C, or T	373	100
498	AAB70537	Homo sapiens	Human PRO7 protein sequence SEQ ID NO:14.	373	100
498	gi13185725	Homo sapiens	n 1755 can be A, G, C, or T.	373	100
499	gi202752	Rattus norvegicus	adenylyl cyclase type II	261	59
499	AAB02006	Homo sapiens	Adenylyl cyclase type II-C2 C2 alpha domain.	261	59
499	gi2204110	Bos taurus	adenylyl cyclase type VII	138	50
500	gi10433645	Homo sapiens	cDNA FLJ12221 fis, clone MAMMA1001091.	1086	69
500	gi10440418	Homo sapiens	mRNA for FLJ00044 protein, partial cds.	1086	69
500	AAB56941	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1519.	126	28
501	AAY99402	Homo sapiens	Human PRO1382 (UNQ718) amino acid sequence SEQ ID NO:220.	492	98

Table 2A

137

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
501	AAY32937	Homo sapiens	Human cerebellin-2 protein sequence.	300	70
501	gi5702371	Mus musculus	precerebellin-1	284	66
502	AAB44681	Homo sapiens	Human secreted protein sequence encoded by gene 41 SEQ ID NO:146.	361	63
502	gi1293734	Saccharomyces cerevisiae	O3635p	279	34
502	gi13877141	Homo sapiens	FKSG89	162	33
503	gi4731216	Boophilus microplus	NADH dehydrogenase subunit 2	52	25
503	gi6180101	Cafeteria roenbergensis	NADH dehydrogenase subunit 2	71	48
503	gi5869819	Globodera pallida	NADH-ubiquinone oxidoreductase subunit 1	82	35
504	AAY34120	Homo sapiens	Human potassium channel K+Hnov4.	1597	99
504	gi206044	Rattus norvegicus	potassium channel Kv3.2b	1582	98
504	gi206914	Rattus norvegicus	K+ channel protein	1582	98
505	gi3790674	Caenorhabditis elegans	contains similarity to a vac1/fab1-type domain	449	54
506	AAB53626	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:1166.	55	47
506	gi1049106	Homo sapiens	Human dystonin isoform 2 mRNA, partial cds.	63	100
506	gi470480	Homo sapiens	Human clone JL8 immunoglobulin kappa chain (IgK) mRNA, VKIII-JK3 region, partial cds.	58	34
507	AAY44985	Homo sapiens	Human epidermal protein-2.	82	37
507	gi11073	Drosophila melanogaster	Mst84Da	75	37
507	gi8571115	Homo sapiens	human endogenous retrovirus HRES-1 p8 protein (p8) and p15 protein (p15) genes, complete cds.	75	40
508	gi13676322	Homo sapiens	chromosome 1 open reading frame 2, clone MGC:1298, mRNA, complete cds.	230	31
508	gi13938585	Homo sapiens	clone MGC:4509, mRNA, complete cds.	230	31
508	gi2564916	Homo sapiens	clk2 kinase (CLK2), propin1, cote1, glucocerebrosidase (GBA), and metaxin genes, complete cds; metaxin pseudogene and glucocerebrosidase pseudogene; and thrombospondin3 (THBS3) gene, partial cds.	229	31
509	gi56463	Rattus norvegicus	gp210 (AA 1-1886)	363	79
509	gi6650678	Mus musculus	nuclear pore membrane glycoprotein POM210	358	78
509	gi1703554	Caenorhabditis elegans	strong similarity to rat integral membrane glycoprotein GP120 precursor (SP:P11654)	143	32

Table 2A

138

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
510	AAB73355	Homo sapiens	Human mesangial cell meg-1 protein.	317	52
510	gi4191594	Homo sapiens	protein serine/threonine phosphatase 4 regulatory subunit 1 (PP4R1) mRNA, complete cds.	292	52
510	gi10120321	Salmo trutta	MHC class II alpha chain	58	30
511	gi11320944	Homo sapiens	peptide deformylase-like protein mRNA, complete cds.	1300	100
511	gi13195254	Homo sapiens	polypeptide deformylase-like protein (PDF) mRNA, complete cds.	1300	100
511	gi11320968	Lycopersicon esculentum	peptide deformylase-like protein	346	40
512	gi13279254	Homo sapiens	Similar to RIKEN cDNA 2610207I16 gene, clone MGC:10940, mRNA, complete cds.	417	94
512	gi5869811	Glomus mosseae	Fox2 protein	187	30
512	gi432977	Homo sapiens	Human sterol carrier protein 2 mRNA, complete cds.	174	32
513	gi10803406	Homo sapiens	mRNA for cadherin-19 (CDH19 gene).	863	100
513	AAY41725	Homo sapiens	Human PRO941 protein sequence.	863	100
513	AAB44281	Homo sapiens	Human PRO941 (UNQ478) protein sequence SEQ ID NO:264.	863	100
514	AAB08944	Homo sapiens	Human secreted protein sequence encoded by gene 19 SEQ ID NO:101.	206	83
514	AAB08909	Homo sapiens	Human secreted protein sequence encoded by gene 19 SEQ ID NO:66.	159	80
514	gi14029247	Gnorimosphaeroma oregonense	cytochrome oxidase subunit I	66	53
515	AAG02731	Homo sapiens	Human secreted protein, SEQ ID NO: 6812.	67	38
515	gi1841964	Toxocara canis	TcH SLdT.460	63	37
515	gi3986598	Ginglymostoma cirratum	antigen receptor	58	47
516	gi575501	Homo sapiens	thyrotropin beta-subunit (TSHB) gene, exon 3.	739	99
516	gi339998	Homo sapiens	Human thyrotropin beta (TSH-beta) subunit gene, exons 2 and 3.	739	99
516	gi340002	Homo sapiens	Human thyrotropin beta subunit gene, exons 2 and 3.	739	99
517	AAB53436	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:976.	368	97
517	AAB25691	Homo sapiens	Human secreted protein sequence encoded by gene 27 SEQ ID NO:80.	168	93
517	AAY01428	Homo sapiens	Secreted protein encoded by	81	42

Table 2A

139

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			gene 46 clone HAQBT52.		
518	AAB54178	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:630.	1025	99
518	gi7321824	Drosophila melanogaster	out at first	510	38
518	gi2443448	Drosophila virilis	out at first	508	39
519	AAW75178	Homo sapiens	Human secreted protein encoded by gene 69 clone HPEBD70.	45	47
519	gi6466876	Kashmir bee virus	RNA polymerase	72	43
519	gi6646671	cloudy wing virus	RNA polymerase	72	43
520	AAB88377	Homo sapiens	Human membrane or secretory protein clone PSEC0113.	379	91
520	gi190506	Homo sapiens	Human PRB1 locus salivary proline-rich protein mRNA, clone cP5, complete cds.	111	32
520	gi190475	Homo sapiens	Human salivary proline-rich protein 1 gene, segment 2.	84	34
521	gi1235645	Cladomyrma cryptata	cytochrome oxidase subunit II	57	50
521	gi4981606	Thermotoga maritima	oligopeptide ABC transporter, permease protein	43	31
521	gi6681644	Yaba monkey tumor virus	similar to vaccinia A14.5L	55	45
522	gi7020918	Homo sapiens	cDNA FLJ20668 fis, clone KAIA585.	461	66
522	AAB54305	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:757.	62	33
522	AAY41352	Homo sapiens	Human secreted protein encoded by gene 45 clone HTXFH55.	58	21
523	AAY54054	Homo sapiens	Angiostatin-binding domain of ABP-1, designated Big-3.	137	39
523	gi9887326	Homo sapiens	angiomotin mRNA, complete cds.	155	37
523	AAY54052	Homo sapiens	An angiogenesis-associated protein which binds plasminogen.	155	37
524	gi11072097	Homo sapiens	MLL/GAS7 fusion protein (MLL/GAS7) mRNA, partial cds.	83	25
524	gi7331837	Caenorhabditis elegans	contains similarity to human X-linked deafness dystonia protein (GB:U66035)	60	25
524	AAG02452	Homo sapiens	Human secreted protein, SEQ ID NO: 6533.	59	44
525	gi13195147	Mus musculus	HCH	953	77
525	gi1339910	Homo sapiens	Human DOCK180 protein mRNA, complete cds.	203	32
525	AAW03515	Homo sapiens	Human DOCK180 protein.	203	32
526	gi854065	Human herpesvirus 6	U88	305	47
526	gi9757150	Leishmania major	extremely cysteine/valine rich protein	284	50
526	gi10434098	Homo sapiens	cDNA FLJ12547 fis, clone NT2RM4000634.	219	38

Table 2A

140

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
527	AAY48278	Homo sapiens	Human prostate cancer-associated protein 64.	98	89
527	AAB58446	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 784.	98	89
527	AAG00214	Homo sapiens	Human secreted protein, SEQ ID NO: 4295.	98	89
529	AAB61421	Homo sapiens	Human TANGO 300 protein.	1583	99
529	AAB23618	Homo sapiens	Human secreted protein SEQ ID NO: 36.	1581	99
529	AAB87592	Homo sapiens	Human PRO1925.	1354	98
530	gi6841194	Homo sapiens	HSPC272	421	66
530	gi12248392	Mus musculus	transcriptional inhibitory factor	90	28
530	gi2853265	Rattus norvegicus	jun dimerization protein 2	90	28
531	gi9964124	Helicobacter pylori	HP0519-like protein	54	45
531	gi6970424	Human papillomavirus type 69	start codon is not identified	59	29
532	gi14330385	Homo sapiens	mRNA for sodium/calcium exchanger, SCL8A3, alternative splice form B (SCL8A3 gene).	178	92
532	gi14330383	Homo sapiens	mRNA for sodium/calcium exchanger SCL8A3, alternative splice form A (SCL8A3 gene).	193	60
532	gi1552526	Rattus norvegicus	sodium-calcium exchanger form 3	178	92
533	gi58028	synthetic construct	suef protein	148	32
533	gi2447210	Paramecium bursaria Chlorella virus 1	a312aR	67	35
534	gi8100892	Human immunodeficiency virus type 1	protease	76	30
534	gi14281259	Human immunodeficiency virus	HIV Protease	71	28
534	gi10504617	Human immunodeficiency virus type 1	protease	71	31
535	gi4128041	Homo sapiens	claudin-9 (CLDN9) gene.	146	37
535	AAB64401	Homo sapiens	Amino acid sequence of human intracellular signalling molecule INTRA33.	146	37
535	gi4325296	Mus musculus	claudin-9	143	36
536	gi10433539	Homo sapiens	cDNA FLJ12133 fis, clone MAMMA1000278.	224	35
536	AAW64461	Homo sapiens	Human secreted protein from clone B121.	218	35
536	gi4406644	Homo sapiens	clone 25130 mRNA sequence, complete cds.	223	41
537	AAY05376	Homo sapiens	Human HCMV inducible gene protein, SEQ ID NO 20.	974	90
537	AAB60496	Homo sapiens	Human cell cycle and proliferation protein CCYPR-44, SEQ ID NO:44.	974	90

Table 2A

141

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
537	gi13879501	Mus musculus	RIKEN cDNA 4933419D20 gene	348	41
538	AAY25451	Homo sapiens	Human secreted protein 2 derived from extended cDNA.	123	53
538	AAY35882	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 19.	123	53
538	AAY66636	Homo sapiens	Membrane-bound protein PRO180.	126	47
539	gi14042279	Homo sapiens	cDNA FLJ14627 fis, clone NT2RP2000289.	208	82
539	AAW78193	Homo sapiens	Human secreted protein encoded by gene 68 clone H2CBJ08.	103	46
540	gi10579884	Halobacterium sp. NRC-1	Vng0244h	68	32
541	AAY19740	Homo sapiens	SEQ ID NO 458 from WO9922243.	60	36
541	gi5911915	Homo sapiens	mRNA; cDNA DKFZp586M0622 (from clone DKFZp586M0622); partial cds.	68	31
541	gi4574260	Haemophilus influenzae	outer membrane protein 26	70	29
542	gi13543049	Mus musculus	Similar to RIKEN cDNA 0610030G03 gene	1147	87
542	gi5263332	Arabidopsis thaliana	F8K7.23	123	24
542	gi6552728	Arabidopsis thaliana	T26F17.1	123	24
543	gi14290586	Homo sapiens	Similar to RIKEN cDNA 2810403L02 gene, clone IMAGE:3868486, mRNA, partial cds.	1809	100
543	gi11493522	Homo sapiens	PRO1512	1512	100
543	AAB58871	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 579.	1412	92
544	gi2114213	Homo sapiens	immunoglobulin lambda gene locus DNA, clone:123E1 upstream contig.	788	100
544	gi2114308	Homo sapiens	immunoglobulin lambda gene locus DNA, clone:123E1.	788	100
544	gi693811	human, chromosome 22, Genomic, 1100 nt]. [Homo sapiens	Vpre-B=VPre-B protein	788	100
545	gi14250299	Homo sapiens	Similar to RIKEN cDNA C030006K11 gene, clone MGC:18180, mRNA, complete cds.	686	87
545	gi7230571	Mus musculus	lim homeodomain-containing transcription factor	87	26
545	gi587461	Mesocricetus auratus	lmx1.1	83	25
546	AAB24074	Homo sapiens	Human PRO1153 protein sequence SEQ ID NO:49.	130	34
546	AAY66735	Homo sapiens	Membrane-bound protein PRO1153.	130	34
546	AAB65258	Homo sapiens	Human PRO1153 (UNQ583) protein sequence SEQ ID	130	34

Table 2A

142

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO:351.		
547	gi1537002	Hepatitis C virus	envelope glycoprotein E2/NS1	61	32
547	gi3153687	Hepatitis C virus	genome polyprotein	60	41
547	AAB45374	Homo sapiens	Human secreted protein sequence encoded by gene 36 SEQ ID NO:126.	58	50
548	gi405956	Escherichia coli	yeeE	1138	93
548	gi405954	Escherichia coli	exonuclease I	1014	86
548	gi1736685	Escherichia coli	Exodeoxyribonuclease I (EC 3.1.11.1) (Exonuclease I) (DNA deoxyribophosphodiesterase) (DRPase).	1014	86
549	gi295196	Salmonella typhimurium	level of amino acid identity between E. coli and S.typhimurium strongly suggests authentic gene	699	86
549	gi405956	Escherichia coli	yeeE	96	36
549	AAG01568	Homo sapiens	Human secreted protein, SEQ ID NO: 5649.	65	25
550	AAW67894	Homo sapiens	Human secreted protein encoded by gene 2 clone HBMCF37.	60	28
550	AAY87145	Homo sapiens	Human secreted protein sequence SEQ ID NO:184.	60	28
550	AAY87182	Homo sapiens	Human secreted protein sequence SEQ ID NO:221.	60	28
551	gi216539	Escherichia coli	BasS	825	98
551	gi1790551	Escherichia coli K12	sensor protein for basR	825	98
551	gi536956	Escherichia coli	basS	825	98
552	gi1786804	Escherichia coli K12	ferric enterobactin transport protein	1021	100
552	gi1778505	Escherichia coli	ferric enterobactin transport protein	1021	100
552	gi13360086	Escherichia coli O157:H7	ferric enterobactin transport protein	1020	99
553	gi349227	Escherichia coli	transmembrane protein	1114	100
553	gi466681	Escherichia coli	dppC	1114	100
553	gi13363896	Escherichia coli O157:H7	dipeptide transport system permease protein 2	1114	100
554	gi4063042	Cryptosporidium parvum	GP900; mucin-like glycoprotein	359	57
554	gi2827460	Cercopithecus aethiops	hepatitis A virus cellular receptor 1 short form	324	56
554	gi2827462	Cercopithecus aethiops	hepatitis A virus cellular receptor 1 long form	324	56
555	gi13959789	Homo sapiens	lung alpha/beta hydrolase protein 1 mRNA, complete cds.	203	88
555	gi13784946	Mus musculus	alpha/beta hydrolase-1	175	77
555	gi7545019	Neurospora crassa	apocytochrome b	47	41
556	AAB87774	Homo sapiens	Human T2R44 amino acid sequence SEQ ID NO:70.	364	91
556	AAB87780	Homo sapiens	Human T2R50 amino acid sequence SEQ ID NO:76.	363	89
556	AAB87745	Homo sapiens	Human T2R15 amino acid sequence SEQ ID NO:28.	343	85

Table 2A

143

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
557	gi2275592	Homo sapiens	T cell receptor beta locus, TCRBV8S5P to TCRBV21S2A2 region.	534	100
557	gi2275570	Homo sapiens	T cell receptor beta locus, TCRBV6S4A1 to TCRBV8S1 region.	534	100
557	gi2218039	Homo sapiens	Human germline T-cell receptor beta chain TCRBV13S1, TCRBV6S8A2T, TCRBV5S6A3N2T, TCRBV13S6A2T, TCRBV6S9P, TCRBV5S3A2T, TCRBV13S8P, TCRBV6S3A1N1T, TCRBV5S2, TCRBV6S6A2T, TCRBV5S7P, TCRBV13S4, TCRBV6S2A1N1T, TCRBV5S4A2T, TCRBV6S4A1, TCRBV23S1A2T, TCRBV12S1A1N2, TCRBV21S2A2, TCRBV8S1, TCRBV8S2A1T, TCRBV8S3, TCRBV16S1A1N1, TCRBV24S1A3T, TCRBV25S1A2PT, TCRBV26S1P, TCRBV18S1, TCRBV17S1A1T, TCRBV2S1, TCRBV10S1P genes from bases 257519 to 472940 (section 2 of 3).	534	100
558	gi3093754	Neurospora crassa	AR2	78	28
558	gi3776090	Mus musculus	wolframin	76	29
558	gi3777585	Mus musculus	transmembrane protein	76	29
559	gi2935614	Homo sapiens	PAC clone RP1-102K2 from 22q12.1-qter, complete sequence.	1306	100
559	gi386988	Homo sapiens	Human oncostatin M gene, exon 3.	1306	100
559	AAR33380	Homo sapiens	Cytokine hOSM.	1306	100
560	AAB49502	Homo sapiens	Clone HYASC03.	310	98
560	gi7020468	Homo sapiens	cDNA FLJ20396 fis, clone KAT00561.	145	39
560	AAB18980	Homo sapiens	Amino acid sequence of a human transmembrane protein.	145	39
561	AAY38432	Homo sapiens	Human secreted protein encoded by gene No. 3.	81	46
561	AAY73420	Homo sapiens	Human secreted protein clone ye22_1 protein sequence SEQ ID NO:62.	75	33
561	AAY20298	Homo sapiens	Human apolipoprotein E mutant protein fragment 11.	77	30
562	gi9948048	Pseudomonas aeruginosa	probable transporter (membrane subunit)	557	63
562	gi7227389	Neisseria	sodium/dicarboxylate symporter	492	58

Table 2A

144

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		meningitidis MC58	family protein		
562	gi9657417	Vibrio cholerae	sodium/dicarboxylate symporter	474	55
563	gi13111711	Homo sapiens	solute carrier family 2 (facilitated glucose transporter), member 5, clone MGC:1619, mRNA, complete cds.	1273	60
563	gi12804761	Homo sapiens	solute carrier family 2 (facilitated glucose transporter), member 5, clone MGC:3654, mRNA, complete cds.	1273	60
563	gi183298	Homo sapiens	Human glucose transport-like 5 (GLUT5) mRNA, complete cds.	1273	60
564	gi14336709	Homo sapiens	16p13.3 sequence section 3 of 8.	358	57
564	gi9621664	Homo sapiens	RHBDL gene for rhomboid-related protein.	358	57
564	gi3287191	Homo sapiens	mRNA for rhomboid-related protein, complete CDS.	358	57
565	AAY45023	Homo sapiens	Human sensory transduction G-protein coupled receptor-B3.	968	100
565	gi13785657	Mus musculus	candidate taste receptor T1R1	786	77
565	gi13785659	Mus musculus	candidate taste receptor T1R2	303	36
566	gi871498	Oryza sativa	DNA binding protein	86	35
566	gi7160630	Bordetella bronchiseptica	pertactin (P.68)	86	39
566	gi9049498	Bordetella bronchiseptica	pertactin	86	39
567	gi5911988	Homo sapiens	mRNA; cDNA DKFZp434H2235 (from clone DKFZp434H2235); partial cds.	164	73
567	gi5262574	Homo sapiens	mRNA; cDNA DKFZp434G173 (from clone DKFZp434G173); complete cds.	164	73
567	AAW89030	Homo sapiens	Polypeptide fragment encoded by gene 165.	147	64
568	gi10437864	Homo sapiens	cDNA: FLJ21709 fis, clone COL10077.	429	74
568	AAY91433	Homo sapiens	Human secreted protein sequence encoded by gene 33 SEQ ID NO:154.	412	76
568	gi14042074	Homo sapiens	cDNA FLJ14508 fis, clone NT2RM1000421, weakly similar to RIBONUCLEASE INHIBITOR.	411	80
569	gi9280561	Mus musculus	elafin-like protein I	66	30
569	AAY99453	Homo sapiens	Human PRO1784 (UNQ846) amino acid sequence SEQ ID NO:390.	77	31
569	gi10176740	Arabidopsis thaliana	RING zinc finger protein-like	76	33
570	AAB87396	Homo sapiens	Human gene 8 encoded secreted protein HMAM121, SEQ ID NO:137.	440	89
570	AAY95967	Homo sapiens	Human TANGO 240.	436	88
570	AAB88402	Homo sapiens	Human membrane or secretory protein clone PSEC0152.	434	88

Table 2A

145

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
571	AAY19485	Homo sapiens	Amino acid sequence of a human secreted protein.	53	52
572	gi6900006	Ceratitis capitata	chorion protein s18	95	31
572	gi1491621	Bovine herpesvirus 1	UL36	104	35
572	gi2653311	Bovine herpesvirus type 1.1	very large virion protein (tegument)	104	35
573	gi4877582	Homo sapiens	lipoma HMGIC fusion partner (LHFP) mRNA, complete cds.	72	34
573	AAY87336	Homo sapiens	Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	72	34
573	gi9658445	Vibrio cholerae	AzlC family protein	49	38
574	gi6899191	Ureaplasma urealyticum	amino acid antiporter	67	33
574	gi5708228	Rhodopseudomonas acidophila	LH2alpha7	62	35
574	gi7211354	Saimiri boliviensis	olfactory receptor	77	34
575	AAB19403	Homo sapiens	Amino acid sequence of a human secreted protein.	712	89
575	gi387048	Cricetus cricetus	DHFR-coamplified protein	230	47
575	gi3261597	Mycobacterium tuberculosis	lprA	77	29
576	gi12718841	Mus musculus	Skullin	310	38
576	gi4191356	Mus musculus	claudin-6	308	38
576	gi13543081	Mus musculus	claudin 6	308	38
577	gi801882	Vibrio alginolyticus	FkuB	83	31
577	gi2795895	Homo sapiens	clone 23819 white protein homolog mRNA, partial cds.	71	30
577	gi5777942	Equus caballus	IL-1ra	52	25
578	gi9872	Plasmodium falciparum	ATPase I	116	41
578	gi7688148	Homo sapiens	Novel human gene mapping to chomosome 1.	119	42
578	gi3451312	Schizosaccharomyces pombe	membrane atpase	116	41
579	gi6682873	Homo sapiens	rec mRNA, complete cds.	200	90
579	gi7230612	Rattus norvegicus	small rec	197	87
579	gi4959442	Drosophila melanogaster	DNZDHHC/NEW1 zinc finger protein 11	93	41
580	gi2204110	Bos taurus	adenylyl cyclase type VII	233	69
580	gi602412	Mus musculus	adenylyl cyclase type VII	209	66
580	AAB02011	Homo sapiens	Type VII adenylyl cyclase.	209	66
581	AAB24476	Homo sapiens	Human secreted protein sequence encoded by gene 40 SEQ ID NO:101.	241	69
581	gi452414	Mus musculus	mPit-1R	69	31
581	gi7769944	Leishmania major	L354.10	87	25
582	gi3297936	Rattus norvegicus	rhomboid-related protein	267	71
582	gi9621664	Homo sapiens	RHBDL gene for rhomboid-related protein.	266	71
582	gi14336709	Homo sapiens	16p13.3 sequence section 3 of 8.	266	71
583	gi10437529	Homo sapiens	cDNA: FLJ21432 fis, clone COL04219.	145	25
583	AAY76136	Homo sapiens	Human secreted protein encoded	113	28

Table 2A

146

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			by gene 13.		
583	gi4929559	Homo sapiens	CGI-45 protein mRNA, complete cds.	113	28
584	gi2429362	Santalum album	proline rich protein	137	34
584	gi5139695	Cucumis sativus	expressed in cucumber hypocotyls	127	28
584	gi7671460	Arabidopsis thaliana	AtAGP4	111	37
585	gi3165565	Caenorhabditis elegans	contains similarity to transmembrane domains found in HMG CoA reductases and drosophila patched protein (SW:P18502)	94	23
585	gi160281	Plasmodium falciparum	erythrocyte binding protein	64	35
585	.AAY28686	Homo sapiens	Human yb39_1 secreted protein.	57	43
587	AAY71948	Homo sapiens	Human ion channel protein (ICP).	1195	99
587	AAY71949	Homo sapiens	Human alternative ion channel protein (ICP).	1195	99
587	AAR27654	Homo sapiens	Human calcium channel 27980/16.	149	27
588	gi478889	Rana catesbeiana	transcription factor RcC/EPB-1	82	33
588	gi4098456	Sus scrofa	follicle-stimulating hormone beta subunit	60	38
588	AAR56767	Homo sapiens	Human FSH beta subunit fragment with residues -18 to 35.	58	33
589	gi5578778	Homo sapiens	mRNA for G18.2 protein (G18.2 gene, located in the class III region of the major histocompatibility complex).	73	41
589	gi213591	Pseudopleuronectes americanus	HPLC6	65	43
589	gi11345434	Thermus thermophilus	competence factor ComEA	79	43
590	gi13111831	Homo sapiens	clone IMAGE:3451448, mRNA, partial cds.	606	60
590	AAW78128	Homo sapiens	Human secreted protein encoded by gene 3 clone HOSBI96.	606	60
590	AAB18993	Homo sapiens	Amino acid sequence of a human transmembrane protein.	606	60
591	gi14249886	Homo sapiens	clone MGC:15763, mRNA, complete cds.	196	77
591	gi217554	Bos taurus	endothelin receptor	50	32
591	gi3299894	Equus caballus	endothelin-B receptor	50	32
592	gi36853	Homo sapiens	Human mRNA for T-cell receptor alpha-chain HAVP02 (V(a)11.1-J(a)1).	585	100
592	gi2358022	Homo sapiens	T-cell receptor alpha delta locus from bases 1 to 250529 (section 1 of 5) of the Complete Nucleotide Sequence.	585	100
592	gi404055	Macaca mulatta	T-cell receptor alpha chain	568	97
593	AAW52812	Homo sapiens	Human induced tumour protein.	123	38
593	gi8895091	Homo sapiens	Diff33 protein homolog mRNA,	123	38

Table 2A

147

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			complete cds.		
593	AAY95015	Homo sapiens	Human secreted protein vc61_1, SEQ ID NO:70.	123	38
594	gi32093	Homo sapiens	H.sapiens HGMP07J gene for olfactory receptor.	849	54
594	AAF61132_aa1	Homo sapiens	Human OLFXY cDNA.	802	49
594	AAB46999	Homo sapiens	Human OLFXY protein.	799	49
595	gi9081843	Prunus dulcis	self-incompatibility associated ribonuclease	79	44
595	gi6539444	Prunus avium	S6-RNase	79	44
595	gi6539438	Prunus avium	S1-RNase	78	44
596	AAB66272	Homo sapiens	Human TANGO 378 SEQ ID NO: 29.	581	100
596	AAB61166	Homo sapiens	Human BBSR seven transmembrane receptor protein.	168	39
596	gi6006811	Mus musculus	serpentine receptor	168	41
597	AAY66750	Homo sapiens	Membrane-bound protein PRO1287.	785	98
597	AAB87561	Homo sapiens	Human PRO1287.	785	98
597	AAB65273	Homo sapiens	Human PRO1287 (UNQ656) protein sequence SEQ ID NO:381.	785	98
598	AAY99421	Homo sapiens	Human PRO1433 (UNQ738) amino acid sequence SEQ ID NO:292.	915	48
598	gi13537297	Homo sapiens	GS1999full mRNA, complete cds.	879	51
598	AAY94889	Homo sapiens	Human protein clone HP02485.	723	43
599	gi10435844	Homo sapiens	cDNA FLJ13737 fis, clone PLACE3000157.	93	28
599	gi205752	Rattus norvegicus	Nopp140	95	27
599	AAY53800	Homo sapiens	Amino acids 145-197 of the mature human chromogranin A (CgA) protein.	63	40
600	gi7717312	Homo sapiens	chromosome 21 segment HS21C049.	422	97
600	AAB18666	Homo sapiens	A human regulator of intracellular phosphorylation.	115	92
600	gi11342496	Bacteriophage phi-Ea1h	holin	77	27
601	gi9963895	Homo sapiens	HT021 (HT021) mRNA, complete cds.	255	94
601	AAW54455	Homo sapiens	Mouse novel secreted protein isolated from clone BF290_li.	255	94
601	AAB59017	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 725.	255	94
602	gi2055228	Glycine max	SRC1	76	26
602	gi204144	Rattus norvegicus	profilaggrin	97	25
602	gi3820941	Hepatitis B virus	core antigen	71	24
603	gi1234787	Xenopus laevis	up-regulated by thyroid hormone in tadpoles; expressed specifically in the tail and only at metamorphosis; membrane	1115	58

Table 2A

148

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			bound or extracellular protein; C-terminal basic region		
603	gi10435980	Homo sapiens	cDNA FLJ13840 fis, clone THYRO1000783, moderately similar to Xenopus laevis tail-specific thyroid hormone up-regulated (gene 5) mRNA.	699	72
603	gi4868122	Mus musculus	hedgehog-interacting protein	405	33
604	gi1181494	Paramecium bursaria Chlorella virus 1	a331L	61	46
604	AAY91469	Homo sapiens	Human secreted protein sequence encoded by gene 19 SEQ ID NO:142.	57	40
604	AAY91617	Homo sapiens	Human secreted protein sequence encoded by gene 19 SEQ ID NO:290.	57	40
605	gi12007419	Mus musculus	B4 olfactory receptor	285	60
605	gi12007420	Mus musculus	B5 olfactory receptor	285	60
605	gi12007421	Mus musculus	B6 olfactory receptor	285	60
606	AAB20695	Homo sapiens	Polymeric immunoglobulin receptor binding domain peptide SEQ ID NO:11.	60	55
606	gi1181346	Paramecium bursaria Chlorella virus 1	a183L	56	28
606	gi14030701	Arabidopsis thaliana	At2g28370/T1B3.11	72	27
607	gi13507259	Homo sapiens	amnionless mRNA, complete cds.	1167	99
607	gi13649780	Mus musculus	amnionless precursor protein	840	71
607	AAY66714	Homo sapiens	Membrane-bound protein PRO1028.	1167	99
609	gi1296632	Homo sapiens	H.sapiens gene encoding G protein coupled receptor.	104	37
609	gi1124905	Homo sapiens	H.sapiens P2Y4 gene.	104	37
609	AAW23606	Homo sapiens	Human P2Y4 receptor polypeptide.	104	37
610	gi4877582	Homo sapiens	lipoma HMGIC fusion partner (LHFP) mRNA, complete cds.	110	25
610	AAY87336	Homo sapiens	Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	110	25
611	AAY27721	Homo sapiens	Human secreted protein encoded by gene No. 29.	1118	88
611	AAB87068	Homo sapiens	Human secreted protein TANGO 365, SEQ ID NO:46.	621	99
611	AAB87146	Homo sapiens	Human secreted protein TANGO 365 A5V variant, SEQ ID NO:161.	617	98
612	gi7208423	Caulobacter crescentus	CpaA	65	36
612	gi13424575	Caulobacter crescentus	pilus assembly protein CpaA	65	36
613	AAY28917	Homo sapiens	Human regulatory protein HRGP-3.	267	100
613	AAB53312	Homo sapiens	Human colon cancer antigen	267	100

Table 2A

149

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			protein sequence SEQ ID NO:852.		
613	gi11526789	Homo sapiens	inorganic pyrophosphatase 2 (PPA2) mRNA, complete cds, nuclear gene for mitochondrial product.	258	98
614	gi13938575	Homo sapiens	Similar to RIKEN cDNA 2610511E22 gene, clone MGC:4251, mRNA, complete cds.	655	89
614	AAY91458	Homo sapiens	Human secreted protein sequence encoded by gene 8 SEQ ID NO:131.	655	89
614	AAY91598	Homo sapiens	Human secreted protein sequence encoded by gene 8 SEQ ID NO:271.	655	89
615	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	1026	82
615	gi3860513	Mus famulus	reverse transcriptase	482	84
615	gi4379237	Mus musculus	reverse transcriptase	477	83
616	gi14190365	Arabidopsis thaliana	AT5g17300/MKP11_15	64	32
616	gi11275913	Protophormia atriceps	cytochrome oxidase subunit 1	55	44
616	AAY29337	Homo sapiens	Human secreted protein clone gg894_13 alternate reading frame protein.	63	28
617	AAY20840	Homo sapiens	Human neurofilament-H wild type protein fragment 1.	67	38
617	gi10584099	Halobacterium sp. NRC-1	Vng6036h	61	28
617	gi7739781	Rattus norvegicus	CCN family protein COP-1	80	26
618	gi13183881	Homo sapiens	Fanconi anemia complementation group D2 protein (FANCD2) mRNA, complete cds, alternatively spliced.	657	90
618	gi13324523	Homo sapiens	Fanconi anemia complementation group D2 protein (FANCD2) gene, exons 43, 44, and complete cds, alternatively spliced.	657	90
618	gi10434106	Homo sapiens	cDNA FLJ12551 fis, clone NT2RM4000700.	175	100
619	gi14042550	Homo sapiens	cDNA FLJ14779 fis, clone NT2RP4000398, moderately similar to ZINC FINGER PROTEIN 140.	242	66
619	gi456269	Mus musculus domesticus	zinc finger protein 30	242	70
619	gi5080758	Homo sapiens	chromosome 19, BAC 331191 (CIT-B-471f3), complete sequence.	244	69
620	AAB47106	Homo sapiens	Second splice variant of MAPP.	223	97
620	AAB47105	Homo sapiens	First splice variant of MAPP.	200	90
620	AAW25722	Homo sapiens	Human partial beta meltrin protein fragment 2.	184	66

Table 2A

150

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
621	AAB90649	Homo sapiens	Human secreted protein, SEQ ID NO: 192.	563	92
621	AAB90565	Homo sapiens	Human secreted protein, SEQ ID NO: 103.	472	100
621	AAB90651	Homo sapiens	Human secreted protein, SEQ ID NO: 194.	203	97
622	AAY87335	Homo sapiens	Human signal peptide containing protein HSPP-112 SEQ ID NO:112.	623	99
622	gi2292988	Rattus norvegicus	Inter-alpha-inhibitor H4 heavy chain	87	32
622	AAY90288	Homo sapiens	Human peptidase, HPEP-5 protein sequence.	63	36
623	AAY92710	Homo sapiens	Human membrane-associated protein Zsig24.	230	100
623	AAY87250	Homo sapiens	Human signal peptide containing protein HSPP-27 SEQ ID NO:27.	230	100
623	AAG00627	Homo sapiens	Human secreted protein, SEQ ID NO: 4708.	93	100
624	gi10441465	Homo sapiens	actin filament associated protein (AFAP) mRNA, complete cds.	274	90
624	gi13129531	Gallus gallus	actin filament-associated protein	204	71
624	gi13129529	Gallus gallus	neural actin filament protein	204	71
625	AAB64802	Homo sapiens	Human secreted protein sequence encoded by gene 30 SEQ ID NO:88.	58	41
625	gi1711217	Caenorhabditis elegans	F58A3.1b	77	30
625	gi1711215	Caenorhabditis elegans	F58A3.1a	77	30
626	AAB12121	Homo sapiens	Hydrophobic domain protein from clone HP02962 isolated from KB cells.	153	68
626	AAY30812	Homo sapiens	Human secreted protein encoded from gene 2.	149	65
626	AAB88452	Homo sapiens	Human membrane or secretory protein clone PSEC0241.	144	66
627	gi13623237	Homo sapiens	clone MGC:10671, mRNA, complete cds.	146	57
627	gi13310191	multiple sclerosis associated retrovirus element	recombinant envelope protein	126	35
627	gi4262296	Homo sapiens	endogenous retrovirus W envelope protein mRNA, partial cds.	117	35
628	gi10437485	Homo sapiens	cDNA: FLJ21394 fis, clone COL03536.	65	30
628	AAG02270	Homo sapiens	Human secreted protein, SEQ ID NO: 6351.	59	44
629	gi4200216	Homo sapiens	H.sapiens gene from PAC 1026E2, partial.	475	100
629	gi14141674	Rattus norvegicus	BMP/retinoic acid-inducible neural-specific protein	151	54

Table 2A

151

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
629	gi3041877	Homo sapiens	IB3089A (IB3089A) mRNA, complete cds.	151	54
630	AAY20292	Homo sapiens	Human apolipoprotein E wild type protein fragment 2.	63	51
630	AAB32406	Homo sapiens	Human secreted protein sequence encoded by gene 5 SEQ ID NO:92.	62	36
630	gi12667610	uncultured sulfate-reducing bacterium UMTRAdsr648-22	dissimilatory sulfite reductase subunit A	72	39
631	gi12053099	Homo sapiens	mRNA; cDNA DKFZp434A171 (from clone DKFZp434A171); complete cds.	172	65
631	gi3002799	Pseudomonas pseudoalcaligenes	2-aminomuconic acid semialdehyde dehydrogenase	118	29
631	gi5821145	Homo sapiens	mRNA for RNA binding protein, partial cds, clone: R11.	120	22
632	gi14249823	Homo sapiens	cholecystokinin, clone MGC:10571, mRNA, complete cds.	356	100
632	gi179996	Homo sapiens	Human cholecystokinin (CCK) gene, exon 3.	356	100
632	AAB24381	Homo sapiens	Human procholecystokinin amino acid sequence SEQ ID NO:1.	356	100
633	gi1870554	Saguinus oedipus	T-cell receptor beta	79	32
633	gi1150925	Bovine herpesvirus 1	glycoprotein B	65	38
633	gi159250	Holothuria tubulosa	sperm specific protein phi-0	60	30
634	gi4097231	Ureaplasma urealyticum	multiple banded antigen	395	23
634	gi560649	Neocallimastix patriciarum, Peptide, 860 aa	Xylanase B, XYLB {EC 3.2.1.8}	330	20
634	gi600118	Zea mays	extensin-like protein	331	35
635	AAB12140	Homo sapiens	Hydrophobic domain protein isolated from WERI-RB cells.	172	51
635	AAY25806	Homo sapiens	Human secreted protein fragment encoded from gene 23.	130	46
635	gi5901846	Drosophila melanogaster	BcDNA.GH12144	124	39
636	AAB66267	Homo sapiens	Human TANGO 272 SEQ ID NO: 14.	1329	97
636	gi2289904	Mus musculus	DRPLA	125	28
636	gi1549217	Mus musculus	DRPLA protein	124	28
637	gi4705	Saccharomyces cerevisiae	Ty protein	58	51
637	gi11139690	Ovis aries	muscle specific calpain 3	54	41
637	AAY41363	Homo sapiens	Human secreted protein encoded by gene 56 clone HNGFE55.	54	55
638	gi13926111	Homo sapiens	2P domain potassium channel Talk-2 (KCNK17) mRNA, complete cds.	1430	100
638	AAY90354	Homo sapiens	Human TWIK-3 protein.	1426	99
638	gi13507377	Homo sapiens	potassium channel TASK-4	1364	99

Table 2A

152

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			mRNA, complete cds.		
639	gi514916	Bos taurus	tau protein	91	36
639	gi437055	Macaca mulatta	mucin	95	28
639	gi2754696	Gallus gallus	high molecular mass nuclear antigen	103	28
640	gi14193307	Candidatus Carsonella ruddii	ATP synthase beta subunit	61	35
640	gi2688677	Borrelia burgdorferi	oligopeptide ABC transporter, permease protein (oppC-2)	65	28
640	gi14193323	Candidatus Carsonella ruddii	ATP synthase beta subunit	59	31
641	gi3127175	Homo sapiens	sulfonylurea receptor 2A (SUR2) gene, alternatively spliced product, exon 38a and complete cds.	713	98
641	gi3127176	Homo sapiens	sulfonylurea receptor 2B (SUR2) gene, alternatively spliced product, exon 38b and complete cds.	713	98
641	gi5814019	Oryctolagus cuniculus	cardiac ventricle sulfonyl urea receptor	678	93
642	AAB24035	Homo sapiens	Human PRO4397 protein sequence SEQ ID NO:42.	1894	100
642	AAY93951	Homo sapiens	Amino acid sequence of a Brainiac-5 polypeptide.	1241	100
642	AAY06462	Homo sapiens	Human Brainiac-3.	553	48
643	AAW88708	Homo sapiens	Secreted protein encoded by gene 175 clone HEMAM41.	747	87
643	gi159655	Ascaris suum	collagen	94	36
643	gi289662	Caenorhabditis elegans	col-36 collagen	109	41
644	gi975893	Homo sapiens	Human apolipoprotein apoC-IV (APOC4) gene, complete cds.	693	100
644	AAG03772	Homo sapiens	Human secreted protein, SEQ ID NO: 7853.	669	96
644	gi1185465	Oryctolagus cuniculus	Apolipoprotein C-IV	379	55
645	AAY57878	Homo sapiens	Human transmembrane protein HTMPN-2.	101	86
645	gi4406500	Carassius auratus	gonadotropin releasing hormone receptor type A	72	31
646	AAY59682	Homo sapiens	Secreted protein 108-009-5-0-A2-FL.	488	100
646	AAY01635	Homo sapiens	Human PS214 derived polypeptide.	488	100
646	AAY64650	Homo sapiens	Human luman homology protein.	488	100
647	gi13442978	Mus musculus	D-glucuronyl C5-epimerase	1001	94
647	gi11935177	Mus musculus	heparin/heparan sulfate:glucuronic acid C5 epimerase	1001	94
647	gi13654639	Bos taurus	D-glucuronyl C5 epimerase	972	92
648	AAG00122	Homo sapiens	Human secreted protein, SEQ ID NO: 4203.	102	100

Table 2A

153

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
648	gi4583535	Homo sapiens	integrin alpha 2 subunit (ITGA2) DNA, 5' UTR and promoter region.	99	95
648	AAW70542	Homo sapiens	Integrin alpha-2 chain.	102	100
649	AAY01387	Homo sapiens	Secreted protein encoded by gene 5 clone HTLFE42.	60	40
649	gi3406819	Mus musculus	growth factor receptor	58	38
649	AAG02139	Homo sapiens	Human secreted protein, SEQ ID NO: 6220.	53	40
650	AAB12150	Homo sapiens	Hydrophobic domain protein isolated from HT-1080 cells.	683	100
650	gi13096862	Mus musculus	RIKEN cDNA 9430096L06 gene	634	90
650	AAB29651	Homo sapiens	Human membrane-associated protein HUMAP-8.	502	100
651	gi14250140	Homo sapiens	clone MGC:14809, mRNA, complete cds.	173	100
651	gi561639	Homo sapiens	IgE receptor beta chain (HTm4) mRNA, complete cds.	173	100
651	AAW06503	Homo sapiens	HTm4 protein.	173	100
652	AAY41428	Homo sapiens	Fragment of human secreted protein encoded by gene 17.	107	43
652	AAY41324	Homo sapiens	Human secreted protein encoded by gene 17 clone HNFIY77.	108	40
652	AAB67576	Homo sapiens	Amino acid sequence of a human hydrolytic enzyme HYENZ8.	108	40
653	gi7209315	Homo sapiens	mRNA for FLJ00007 protein, partial cds.	1024	79
653	AAY99428	Homo sapiens	Human PRO1431 (UNQ737) amino acid sequence SEQ ID NO:315.	430	93
653	gi6599145	Homo sapiens	mRNA; cDNA DKFZp434L127 (from clone DKFZp434L127); partial cds.	320	33
654	gi297172	Rattus rattus	ribosomal protein S7	432	93
654	gi2811284	Mus musculus	ribosomal protein S7	432	93
654	gi12804027	Homo sapiens	ribosomal protein S7, clone MGC:10268, mRNA, complete cds.	432	93
655	AAB68888	Homo sapiens	Human RECAP polypeptide, SEQ ID NO: 18.	277	64
655	AAB08944	Homo sapiens	Human secreted protein sequence encoded by gene 19 SEQ ID NO:101.	74	72
655	AAY76198	Homo sapiens	Human secreted protein encoded by gene 75.	67	59
656	gi4096055	Homo sapiens	chromosome 19, cosmid R28379, complete sequence.	136	100
656	gi9950071	Pseudomonas aeruginosa	probable permease of ABC transporter	81	39
656	gi2113989	Mycobacterium tuberculosis	ccsA	79	34
657	gi10438804	Homo sapiens	cDNA: FLJ22419 fis, clone	262	92

Table 2A

154

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			HRC08593.		
657	gi10436785	Homo sapiens	cDNA FLJ14342 fis, clone THYRO1000569, highly similar to Mus musculus hematopoietic zinc finger protein mRNA.	98	42
657	gi6690339	Mus musculus	hematopoietic zinc finger protein	96	40
658	gi9963845	Homo sapiens	HT017 mRNA, complete cds.	558	38
658	AAW09405	Homo sapiens	Pineal gland specific gene-1 protein.	558	38
658	AAB69185	Homo sapiens	Human hISLR-iso protein SEQ ID NO:7.	558	38
659	gi475542	Rattus norvegicus	glutamate receptor delta-1 subunit	505	98
659	gi220418	Mus musculus	glutamate receptor channel subunit delta-1	505	98
659	gi56286	Rattus norvegicus	glutamate receptor subtype delta-1	482	98
660	AAB61880	Homo sapiens	Human cytokine receptor Zcytor14.	163	28
660	AAB61881	Homo sapiens	Human variant Zcytor14 protein Zcytor14-1.	137	32
660	AAB87606	Homo sapiens	Human PRO20040.	143	28
661	gi13195147	Mus musculus	HCH	413	86
661	gi1339910	Homo sapiens	Human DOCK180 protein mRNA, complete cds.	373	78
661	AAW03515	Homo sapiens	Human DOCK180 protein.	366	76
662	AAY27669	Homo sapiens	Human secreted protein encoded by gene No. 103.	255	100
662	gi3719255	Mus musculus	C1q/MBL/SPA receptor C1qRp	50	35
662	gi5714405	Mus musculus	C1q/MBL/SP-A phagocytic receptor C1qRp	50	35
663	gi12724402	Lactococcus lactis subsp. lactis	prophage pi3 protein 41	58	36
663	gi155287	Vibrio cholerae	disulfide isomerase	73	29
664	gi6822060	Arabidopsis thaliana	peptide transport-like protein	93	31
664	gi206311	Rattus norvegicus	protein phosphatase-2Bc	58	30
665	gi14042519	Homo sapiens	cDNA FLJ14763 fis, clone NT2RP3003621.	2026	99
665	gi13097630	Homo sapiens	clone MGC:10791, mRNA, complete cds.	2026	99
665	gi13591620	Homo sapiens	kremen mRNA for kringle-containing transmembrane protein, complete cds.	860	49
666	gi13161409	Mus musculus	family 4 cytochrome P450	437	73
666	gi7331756	Caenorhabditis elegans	contains similarity to Pfam family PF00067 (Cytochrome P450), score=356.1, E=3.6e-103, N=1	139	37
666	gi3876203	Caenorhabditis elegans	contains similarity to Pfam domain: PF00067 (Cytochrome P450), Score=347.4, E-value=5.1e-101, N=1	135	37
667	AAB08862	Homo sapiens	Amino acid sequence of a	958	100

Table 2A

155

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			human secretory protein.		
667	gi12654587	Homo sapiens	clone MGC:2463, mRNA, complete cds.	953	99
667	AAB12163	Homo sapiens	Hydrophobic domain protein from clone HP10671 isolated from Thymus cells.	953	99
668	gi4877582	Homo sapiens	lipoma HMGIC fusion partner (LHFP) mRNA, complete cds.	195	30
668	AAY87336	Homo sapiens	Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	195	30
668	gi7529641	Schizosaccharomyces pombe	calcium permease family membrane transporter	110	28
669	gi3598974	Rattus norvegicus	protein tyrosine phosphatase TD14	105	38
669	gi6625751	Mink enteritis virus	capsid protein VP2	50	34
669	gi5442034	Mus musculus	calmodulin-dependent protein kinase II beta M isoform	66	37
670	AAB33892	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 107.	43	60
670	AAB54248	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:700.	62	42
670	gi683548	Chironomus pallidivittatus	gamma protein constant region	62	38
671	gi41077	Escherichia coli	cal protein precursor (aa 1-51)	63	42
671	gi2995968	Leontopithecus rosalia	NADH dehydrogenase subunit 4	76	28
671	gi2995972	Leontopithecus chrysomelas	NADH dehydrogenase subunit 4	76	28
672	gi1196439	Homo sapiens	(clone H 4.4) latent transforming growth factor-beta binding protein (LTBP-1L) gene, partial cds.	291	98
672	gi207286	Rattus norvegicus	TGF-beta masking protein large subunit	226	77
672	gi3493176	Mus musculus	latent TGF beta binding protein	217	73

Table 2B

156

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
337	AAG81442	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:402.	844	100
337	AAO12909	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26801.	526	100
337	gi12580867	Picea abies	60S ribosomal protein L13E	80	33
338	gi8953907	Mus musculus	thymic stromal lymphopoietin receptor	79	30
338	AAY57951	Homo sapiens	INCY- Human transmembrane protein HTMPN-75.	76	33
338	gi7243288	Mus musculus	cytokine receptor like molecule 2	75	29
339	AAR91305	Homo sapiens	SAKA Transcription factor-IIIA.	96	45
339	gi1616942	Homo sapiens	Xenopus transcription factor IIIA homologue	96	45
339	gi7417372	Homo sapiens	intracellular hyaluronan-binding protein	93	40
340	AAE14342	Homo sapiens	INCY- Human protease PRTS-7 protein.	251	100
340	AAB08950	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 22 SEQ ID NO:107.	251	100
340	AAB08912	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 22 SEQ ID NO:69.	251	100
341	gi|188672|gb|AAA59866.1|	Homo sapiens	mannose 6-phosphate receptor	65	44
341	gi|7025416|gb|AAF35878.1|AF224071_1	Solanum campechiense	NADH dehydrogenase subunit	65	31
342	AAY02361	Homo sapiens	ONOY Polypeptide identified by the signal sequence trap method.	1131	100
342	AAY17526	Homo sapiens	GEMY Human secreted protein clone AM349 2 protein.	1131	100
342	gi20988438	Homo sapiens	Similar to chondroitin beta1,4 N-acetylgalactosaminyltransferase	1131	100
343	gi7768740	Homo sapiens	similar to zinc finger 5 protein	79	29
343	gi20809693	Homo sapiens	Similar to RIKEN cDNA 4933432E21 gene	73	32
343	gi14329696	Homo sapiens	Doublesex-mab-3 (DM) domain	73	32
344	ABB85001	Homo sapiens	GETH Human PRO28631 protein sequence SEQ ID NO:370.	1576	99
344	AAY86234	Homo sapiens	HUMA- Human secreted protein HNTNC20, SEQ ID NO:149.	475	60
344	AAB65258	Homo sapiens	GETH Human PRO1153 (UNQ583) protein sequence SEQ ID NO:351.	109	30
345	gi20072551	Mus musculus	RIKEN cDNA 4930511J11 gene	431	45
345	gi12836893	Gallus gallus	IPR328-like protein	151	30
345	gi17974542	Homo sapiens	voltage-dependent calcium channel gamma-8 subunit	150	26
346	AAA54097_aa1	Homo sapiens	GETH PRO228 cDNA.	396	100
346	AAE17037	Homo sapiens	MILL- Human G protein-	396	100

Table 2B

157

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			coupled receptor, SLGP 7 transmembrane receptor profile.		
346	AAE17031	Homo sapiens	MILL- Human G protein-coupled receptor (GPCR), SLGP.	396	100
347	gi1504002	Homo sapiens	similar to a human major CRK-binding protein DOCK180.	252	100
347	gi13195147	Mus musculus	HCH	206	77
347	AAW03515	Homo sapiens	SHKJ Human DOCK180 protein.	95	43
348	gi|21291633|gb|EAA03778.1|	Anopheles gambiae str. PEST	agCP14673	73	25
349	gi16359249	Mus musculus	RIKEN cDNA 1300010M03 gene	1854	91
349	AAU28182	Homo sapiens	HYSE- Novel human secretory protein, Seq ID No 351.	574	38
349	ABB89832	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2208.	522	39
350	ABB11722	Homo sapiens	HYSE- Human V_segment homologue, SEQ ID NO:2092.	856	99
350	gi1552496	Homo sapiens	V_segment translation product	614	100
350	AAR26977	Homo sapiens	ROUS Human T lymphocyte receptor V-beta 9 subfamily segment.	609	100
351	AAU20502	Homo sapiens	HUMA- Human secreted protein, Seq ID No 494.	162	80
351	gi13960126	Homo sapiens	Similar to leucine-rich neuronal protein	162	80
351	AAU20424	Homo sapiens	HUMA- Human secreted protein, Seq ID No 416.	133	64
352	AAB61141	Homo sapiens	CURA- Human NOV11 protein.	370	86
352	AAU00392	Homo sapiens	CURA- Human secreted protein, POLY4.	370	86
352	AAU08681	Homo sapiens	CURA- Human FCTR3f polypeptide sequence.	370	86
353	AAE01313	Homo sapiens	HUMA- Human gene 2 encoded secreted protein fragment, SEQ ID NO:178.	499	69
353	AAE01233	Homo sapiens	HUMA- Human gene 2 encoded secreted protein HMVAV54, SEQ ID NO:95.	482	69
353	AAE01259	Homo sapiens	HUMA- Human gene 2 encoded secreted protein HMVAV54, SEQ ID NO:121.	476	68
354	AAM95682	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4340.	254	72
354	AAU16249	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1202.	224	95
354	ABB06198	Homo sapiens	BIOW- Human DNA methylation protein 13 SEQ ID NO:2.	193	78
355	AAE07054	Homo sapiens	HUMA- Human gene 4 encoded	680	82

Table 2B

158

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			secreted protein HSYAB05, SEQ ID NO:71.		
355	AAE07077	Homo sapiens	HUMA- Human gene 4 encoded secreted protein HSYAB05, SEQ ID NO:94.	608	76
355	ABB89204	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1580.	456	73
356	AAU91320	Homo sapiens	CYTO- Human P450TEC protein.	865	100
356	gi15080572	Homo sapiens	Similar to RIKEN cDNA 8430436A10 gene	859	100
356	AAE05183	Homo sapiens	INCY- Human drug metabolising enzyme (DME-14) protein.	168	35
357	AAU81988	Homo sapiens	INCY- Human secreted protein SECP14.	484	66
357	AAE06581	Homo sapiens	SAGA Human protein having hydrophobic domain, HP03727.	484	66
357	AAM41951	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6882.	181	94
358	AAE03888	Homo sapiens	HUMA- Human gene 19 encoded secreted protein fragment, SEQ ID NO:140.	359	95
358	AAE03836	Homo sapiens	HUMA- Human gene 19 encoded secreted protein HOGCE48, SEQ ID NO: 82.	359	95
358	ABB11587	Homo sapiens	HYSE- Human peroxidasin homologue, SEQ ID NO:1957.	359	95
359	AAM77193	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 37499.	112	56
359	AAM64370	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36475.	112	56
359	gi7380324	Neisseria meningitidis Z2491	ClpB protein	83	32
360	AAE06576	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10764.	1041	79
360	AAB65258	Homo sapiens	GETH Human PRO1153 (UNQ583) protein sequence SEQ ID NO:351.	1038	79
360	AAG81325	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:168.	1038	79
361	gi15919295	human herpesvirus 5	UL97 protein	71	34
361	gi221797	Human herpesvirus 4	LMP1	70	31
361	gi22938	Human herpesvirus 4	latent membrane protein LMP1	70	31
362	gi3127176	Homo sapiens	sulfonylurea receptor 2B	886	67
362	gi3127175	Homo sapiens	sulfonylurea receptor 2A	886	67
362	gi15778680	Oryctolagus cuniculus	sulphonylurea receptor 2B	873	66

Table 2B

159

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
364	gi18077667	Homo sapiens	bA115P16.2 (inositol 1,4,5-trisphosphate 3-kinase B)	88	32
364	gi14329672	Homo sapiens	inositol 1,4,5-trisphosphate 3-kinase, isoform B	88	32
364	AAE04364	Homo sapiens	INCY- Human kinase (PKIN)-5.	85	32
365	ABB89967	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2343.	462	95
365	AAV42697_aa 1	Homo sapiens	SIBI- DNA encoding human calcium channel alpha-1D subunit.	114	27
365	AAQ84653_aa 1	Homo sapiens	SALK Human neuronal calcium channel subunit alpha 1D.	114	27
366	gi13623421	Homo sapiens	Similar to RIKEN cDNA 5730589L02 gene	571	73
366	gi19484086	Mus musculus	RIKEN cDNA 5730589L02 gene	543	69
366	gi3875896	Caenorhabditis elegans	weak similarity to chalcone flavone isomerase (Swiss Prot accession number P11651)	118	28
367	AAE18208	Homo sapiens	CURA- Human MOL1b protein.	125	87
367	AAY06816	Homo sapiens	UYYA Human Notch2 (humN2) protein sequence.	125	87
367	gi11275978	Homo sapiens	NOTCH 2	125	87
368	AAB47275	Homo sapiens	META- hOAT4.	652	99
368	ABB11750	Homo sapiens	HYSE- Human integral membrane transport protein homologue, SEQ ID NO:2120.	646	98
368	gi18148873	Homo sapiens	hUST3	645	98
369	gi1665787	Homo sapiens	Similar to a C.elegans protein encoded in cosmid C52E12 (U50135)	256	100
369	gi11463949	Homo sapiens	UDP-glucuronic acid	256	100
369	gi14971008	Drosophila melanogaster	UDP-sugar transporter	195	71
370	AAW61626	Homo sapiens	HUMA- Clone HUVBB80 of TM4SF superfamily.	75	26
370	gi15680044	Homo sapiens	Similar to transmembrane 4 superfamily member 1	75	26
370	AAW80948	Homo sapiens	INCY- Amino acid sequence of the human integral membrane protein-2.	73	26
371	ABB06152	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:244.	905	94
371	AAB88377	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0113.	370	94
371	gi15291323	Drosophila melanogaster	GH15686p	315	36
372	AAM40758	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5689.	80	34
373	AAW03515	Homo sapiens	SHKJ Human DOCK180 protein.	120	54
373	gi1339910	Homo sapiens	DOCK180 protein	120	54
373	AAM90486	Homo sapiens	HUMA- Human	118	95

Table 2B

160

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			immune/haematopoietic antigen SEQ ID NO:18079.		
374	AAG77172	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:7938.	215	72
374	gi17946183	Drosophila melanogaster	RE56564p	131	37
374	gi16182326	Drosophila melanogaster	GH01206p	116	20
375	AAB71871	Homo sapiens	MILL- Human GLRP seven transmembrane domain.	73	30
375	AAR70006	Homo sapiens	MERI Human glucagon-like 1 peptide (GLP-1) receptor.	73	30
375	gi717034	Homo sapiens	glucagon-like peptide-1 receptor	73	30
376	gi14189735	Homo sapiens	ATP-binding cassette transporter family A member 12	251	43
376	gi14209836	Mus musculus	ATP-binding cassette transporter sub-family A member 7	199	39
376	AAU09174	Homo sapiens	MILL- Human transporter molecule, MTP-1.	196	40
377	AAM92700	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 2049.	208	67
377	AAB60501	Homo sapiens	INCY- Human cell cycle and proliferation protein CCYPR-49, SEQ ID NO:49.	74	27
377	AAM40936	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5867.	74	27
378	AAY30817	Homo sapiens	HUMA- Human secreted protein encoded from gene 7.	569	98
378	gi3184264	Homo sapiens	F02569_2	101	29
378	gi3386544	Mus musculus	IER5	98	37
379	AAU83223	Homo sapiens	ZYMO Novel secreted protein Z930582G14P.	1440	100
379	AAU83150	Homo sapiens	ZYMO Novel secreted protein Z849065G4P.	1440	100
379	ABB84889	Homo sapiens	GETH Human PRO1415 protein sequence SEQ ID NO:146.	1435	99
380	AAU19385	Homo sapiens	PHAA Human G protein-coupled receptor nGPCR-2318.	219	95
380	gi6636340	Rattus norvegicus	myosin heavy chain Myr 8	157	61
380	gi10863773	Rattus norvegicus	myosin heavy chain Myr 8b	157	61
381	gi18256029	Mus musculus	Similar to RIKEN cDNA 6720456116 gene	270	85
381	gi20988563	Homo sapiens	similar to claudin 19	97	36
381	gi20148965	Mus musculus	claudin 19	97	36
382	gi1679584	Cavia porcellus	membrane cofactor protein precursor	77	37
382	gi1655471	Cavia porcellus	membrane cofactor protein(GMP1-full)	77	37
382	AAV27592_aa 1	Homo sapiens	IMMV Human interleukin-17 receptor cDNA.	73	31

Table 2B

161

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
383	gi2764507	Locusta migratoria	nicotinic acetylcholine receptor, alpha1 subunit	158	38
383	gi9886085	Mus musculus	nicotinic acetlycholine receptor alpha 4 subunit	155	46
383	gi14330017	Mus musculus	bM401L17.2.2 (cholinergic receptor, nicotinic, alpha polypeptide 4 (isoform 2))	155	46
384	gi4995986	Human herpesvirus 6	13.6% identical to DR8 gene of strain U1102 of HHV-6	134	41
384	gi409995	Rattus sp.	mucin	129	42
384	AAM65950	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26256.	123	44
385	ABB06082	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:174.	870	99
385	AAY58174	Homo sapiens	INCY- Human embryogenesis protein, EMPRO.	870	99
385	AAB94377	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14922.	664	73
386	gi13359817	Escherichia coli O157:H7	high-affinity choline transport	1021	100
386	gi1657512	Escherichia coli	high-affinity choline transport protein	1021	100
386	gi12513126	Escherichia coli O157:H7 EDL933	high-affinity choline transport	1021	100
387	gi10584473	Halobacterium sp. NRC-1	Vng6455c	79	27
387	gi10584129	Halobacterium sp. NRC-1	Vng6071c	79	27
387	gi12721708	Pasteurella multocida	UhpB	78	19
388	gi13364609	Escherichia coli O157:H7	fumarate reductase FrdD	515	96
388	gi145266	Escherichia coli	g13 protein	515	96
388	gi12519135	Escherichia coli O157:H7 EDL933	fumarate reductase, anaerobic, membrane anchor polypeptide	515	96
389	gi13363448	Escherichia coli O157:H7	transport protein of hexuronates	928	96
389	gi1160319	Escherichia coli	aldohexuronate transport system	928	96
389	gi12517683	Escherichia coli O157:H7 EDL933	transport of hexuronates	928	96
390	gi395270	Escherichia coli	FepE	402	100
390	gi1778503	Escherichia coli	ferric enterobactin transport protein	402	100
390	gi1786802	Escherichia coli K12	ferric enterobactin (enterochelin) transport	402	100
391	gi13362064	Escherichia coli O157:H7	methyl-accepting chemotaxis protein II	648	83

Table 2B

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
391	gi1736545	Escherichia coli	Methyl-accepting chemotaxis protein II (MCP-II) (Aspartate chemoreceptor protein).	648	83
391	gi145521	Escherichia coli	methyl-accepting chemotaxis protein II	648	83
392	AAM72391	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 32697.	307	100
392	AAM59804	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 31909.	307	100
392	AAB37990	Homo sapiens	HUMA- Human secreted protein encoded by gene 7 clone HWLHH15.	303	98
393	gi3282259	Cucumaria pseudocurata	ND4L	68	30
393	gi|20876844|ref|XP_127831.1|	Mus musculus	similar to ring finger protein 26	68	26
393	gi|3282259|gb|AAC69448.1|	Cucumaria pseudocurata	ND4L	68	30
394	gi13881068	Mycobacterium tuberculosis CDC1551	sugar transporter family protein	83	26
394	gi15074628	Sinorhizobium meliloti	PUTATIVE TRANSMEMBRANE PROTEIN	82	26
394	gi15723037	Burkholderia cepacia	multidrug efflux protein	81	26
395	AAY90272	Homo sapiens	LUDW- Human PTPL1 phosphatase.	81	34
395	AAB19343	Homo sapiens	ISIS- Amino acid sequence of a human Fap-1 (Fas associated protein 1).	81	34
395	AAW75999	Homo sapiens	LUDW- Intracellular protein tyrosine phosphatase, PTPL1.	81	34
396	AAM65947	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26253.	215	25
396	AAM53564	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25669.	215	25
396	gi16412587	Listeria innocua	similar to bacteriophage minor tail proteins	123	14
397	AAO02567	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16459.	351	94
397	AAB88433	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0210.	299	55
397	AAB95155	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17188.	299	55
398	gi1655432	Mus musculus	plexin 2	211	32
398	AAB80241	Homo sapiens	GETH Human PRO235 protein.	208	62
398	AAU12337	Homo sapiens	GETH Human PRO235	208	62

Table 2B

163

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			polypeptide sequence.		
399	AAU81997	Homo sapiens	INCY- Human secreted protein SECP23.	573	100
399	ABB94017	Homo sapiens	HUMA- Human secreted protein SEQ ID NO: 60.	573	100
399	AAB95289	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17509.	573	100
400	AAZ09920_aa 1	Homo sapiens	FARB Human islet cell antigen clone ICA-525 cDNA.	241	40
400	AAV63558_aa 1	Homo sapiens	FARB Islet cell antibody antigen cDNA from clone ICA-525.	241	40
400	AAB48573	Homo sapiens	LUDW- Human breast cancer MO-BC-416 polypeptide.	241	40
401	AAY87340	Homo sapiens	INCY- Human signal peptide containing protein HSPP-117 SEQ ID NO:117.	2104	100
401	gi13543949	Homo sapiens	Similar to RIKEN cDNA 2810432L12 gene	2104	100
401	gi15489421	Mus musculus	RIKEN cDNA 2810432L12 gene	2083	98
402	gi5001993	Dissostichus mawsoni	chimeric AFGP/trypsinogen-like serine protease precursor	195	46
402	gi295736	Dictyostelium discoideum	spore coat protein sp96	186	48
402	gi19570090	Dictyostelium discoideum	Spore coat protein SP96.	186	48
403	gi4206769	Acanthamoeb a castellanii	myosin I heavy chain kinase	131	26
403	gi3599478	Acanthamoeb a castellanii	Myosin-IA	127	34
403	gi2723935	Turnip yellow mosaic virus	No definition line found	116	29
404	AAF90612_aa 1	Homo sapiens	ZYMO Human secretin-like receptor Zgpr1 cDNA.	663	100
404	AAE15635	Homo sapiens	INCY- Human G-protein coupled receptor-5 (GCREC-5) protein.	663	100
404	AAB66272	Homo sapiens	MILL- Human TANGO 378 SEQ ID NO: 29.	663	100
405	gi3850044	Homo sapiens	beta-tubulin cofactor D	94	87
405	gi13111855	Homo sapiens	tubulin-specific chaperone d	94	87
405	gi1465770	Bos taurus	cofactor D	89	75
406	AAC84384_aa 1	Homo sapiens	MILL- Human A236 polypeptide coding sequence.	692	100
406	AAU83656	Homo sapiens	GETH Human PRO protein, Seq ID No 130.	692	100
406	ABB84848	Homo sapiens	GETH Human PRO363 protein sequence SEQ ID NO:64.	692	100
407	AAH77291_aa 1	Homo sapiens	MILL- Human ion channel protein IC23949 cDNA coding region.	791	99
407	AAG77968	Homo sapiens	MILL- Human ion channel protein IC23949.	791	99
407	AAO14211	Homo sapiens	INCY- Human transporter and	791	99

Table 2B

164

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			ion channel TRICH-28.		
408	AAM40199	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3344.	142	76
408	AAM40198	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3343.	142	76
408	AAM41986	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6917.	141	100
409	AAM69908	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30214.	201	100
409	AAM57504	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29609.	201	100
409	gi172177	Saccharomyces cerevisiae	protein kinase C-like protein (PKC1)	81	23
410	AAY99420	Homo sapiens	GETH Human PRO1486 (UNQ755) amino acid sequence SEQ ID NO:287.	1082	100
410	ABB50515	Homo sapiens	HUMA- Human secreted protein encoded by gene 45 SEQ ID NO:463.	1069	99
410	AAW88747	Homo sapiens	HUMA- Secreted protein encoded by gene 45 clone HCESF40.	1069	99
411	AAM93655	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3524.	621	59
411	AAO14195	Homo sapiens	INCY- Human transporter and ion channel TRICH-12.	303	32
411	AAE06584	Homo sapiens	SAGA Human protein having hydrophobic domain, HP03913.	303	32
412	AAE14336	Homo sapiens	INCY- Human protease PRTS-1 protein.	554	100
412	AAB65168	Homo sapiens	GETH Human PRO1310 protein sequence SEQ ID NO:62.	554	100
412	AAU12367	Homo sapiens	GETH Human PRO1310 polypeptide sequence.	554	100
413	gi14794894	Streptomyces nodosus	AmphJ	73	28
413	gi|20833284|ref|XP_131474.1 |	Mus musculus	RIKEN cDNA 9130404H11	185	97
413	gi|14794894|gb|AAK73502.1 |AF357202_5	Streptomyces nodosus	AmphJ	73	28
414	AAM80242	Homo sapiens	HYSE- Human protein SEQ ID NO 3888.	206	92
414	AAM79258	Homo sapiens	HYSE- Human protein SEQ ID NO 1920.	206	92
414	gi5901822	Drosophila melanogaster	EG:118B3.2	160	70
415	gi1834503	Homo sapiens	mucin MUC5B	72	38
416	gi3047402	Homo sapiens	monocarboxylate transporter 2	524	32
416	gi21265165	Homo sapiens	solute carrier family 16 (monocarboxylic acid	523	32

Table 2B

165

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			transporters), member 7		
416	gi2198807	Gallus gallus	monocarboxylate transporter 3	522	34
417	gi6136782	Mus musculus	synaptotagmin V	595	91
417	gi14210264	Rattus norvegicus	synaptotagmin 5	592	91
417	gi1932801	Rattus norvegicus	synaptotagmin X	263	45
418	AAM93692	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3602.	493	100
418	AAB53400	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:940.	493	100
418	ABB89424	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1800.	489	100
419	AAY57952	Homo sapiens	INCY- Human transmembrane protein HTMPN-76.	1142	100
419	AAB24036	Homo sapiens	GETH Human PRO4407 protein sequence SEQ ID NO:47.	1142	100
419	AAB12136	Homo sapiens	PROT- Hydrophobic domain protein from clone HP10625 isolated from Liver cells.	1142	100
420	gi|17532405|ref|NP_495405.1|	Caenorhabditis elegans	C44B7.6.p	72	32
421	AAO14215	Homo sapiens	INCY- Human transporter and ion channel TRICH-32.	213	73
421	AAB47276	Homo sapiens	META- hOAT5.	213	73
421	AAO14213	Homo sapiens	INCY- Human transporter and ion channel TRICH-30.	136	57
422	gi17829	Brassica napus	LEA76 peptide (AA 1-280)	124	26
422	gi13421492	Caulobacter crescentus CB15	methyl-accepting chemotaxis protein McpC	118	20
422	gi20126722	Brassica napus	late embryogenesis-abundant protein	116	25
424	gi13959739	Caprine arthritis-encephalitis virus	envelope glycoprotein	81	33
424	gi323299	Caprine arthritis-encephalitis virus	envelope polyprotein	77	32
424	gi15042572	Ovine lentivirus	variant envelope glycoprotein precursor	77	30
425	ABB89424	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1800.	220	91
425	AAM93692	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3602.	220	91
425	AAB53400	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:940.	220	91
426	AAG72312	Homo sapiens	YEDA Human olfactory	868	92

Table 2B

166

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			receptor polypeptide, SEQ ID NO: 1993.		
426	AAU24606	Homo sapiens	SENO- Human olfactory receptor AOLFR97.	868	92
426	gi18480638	Mus musculus	olfactory receptor MOR205-1	765	80
427	AAG75482	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6246.	90	66
427	AAM67622	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27928.	76	57
427	AAM55226	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 27331.	76	57
428	gi8918871	YccA of plasmid ColIb-P9] [Plasmid F	96 pct identical to gp:AB021078_30	288	98
428	gi|7524597|ref|NP_042351.1|	Pinus thunbergii	protochlorophyllide reductase 58kDa chain	70	37
428	gi|16330680|ref|NP_441408.1|	Synechocystis sp. PCC 6803	ATP synthase e subunit	69	45
429	AAM79503	Homo sapiens	HYSE- Human protein SEQ ID NO 3149.	81	41
429	AAM78519	Homo sapiens	HYSE- Human protein SEQ ID NO 1181.	81	41
429	AAW40058	Homo sapiens	USSH Cellular transcriptional factor CBP.	81	29
430	AAB18985	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	284	31
430	AAE00330	Homo sapiens	ZYMO Human membrane-bound protein-60 (Zsig60).	279	31
430	gi6013381	Rattus norvegicus	TM6P1	279	30
431	AAU16923	Homo sapiens	HUMA- Human novel secreted protein, SEQ ID 164.	346	94
431	gi1934847	Caenorhabditis elegans	DNA topoisomerase I	79	33
431	gi|1934847|emb|CAA65537.1|	Caenorhabditis elegans	DNA topoisomerase; DNA topoisomerase I	79	33
432	gi1913791	Plasmodium vivax	merozoite surface protein	85	30
432	gi537916	Lilium longiflorum	meiotin-1	84	32
432	gi2213848	Plasmodium vivax	merozite surface protein 1	82	30
433	gi17429346	Ralstonia solanacearum	PUTATIVE LIPOPROTEIN	71	36
434	AAU77226	Homo sapiens	DAMB/ Human NR2A N-methyl D-aspartate (NMDA) receptor protein sequence.	159	100
434	AAR80970	Homo sapiens	ALLX Human excitatory amino	159	100

Table 2B

167

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			acid receptor modulatory protein NR2A-1.		
434	AAR55529	Homo sapiens	MERI Human NMDA R2A receptor subunit.	159	100
435	gi18044366	Homo sapiens	Similar to MEGF10 protein	1166	90
435	AAG75479	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6243.	817	62
435	AAB66267	Homo sapiens	MILL- Human TANGO 272 SEQ ID NO: 14.	695	50
436	gi3130157	Takifugu rubripes	pheromone receptor	106	34
436	gi2589210	Mus musculus	calcium-sensing receptor related protein 3	105	35
436	gi2589208	Mus musculus	calcium-sensing receptor related protein 2	99	33
437	gi16605472	Homo sapiens	acyl-malonyl condensing enzyme	1074	99
437	gi4633135	Mus musculus	condensing enzyme	679	50
437	gi2384746	Mus musculus	testicular condensing enzyme	679	50
438	AAG81254	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:26.	1195	86
438	gi7981261	Homo sapiens	dJ50O24.4 (novel protein with DHHC zinc finger domain)	1195	86
438	AAG74779	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5543.	882	64
439	AAO12277	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26169.	74	44
439	gi2209081	Rhytidoponera sp.	cytochrome b	73	25
440	gi12314108	Homo sapiens	dJ230I3.1 (novel protein)	868	84
440	AAB94417	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15016.	578	55
440	gi16416385	Arabidopsis thaliana	anthocyanin-related membrane protein 2	330	33
441	gi20988467	Mus musculus	similar to LD47277p	1185	88
441	AAU91305	Homo sapiens	CORT- Human protein NOV10c.	419	94
441	AAU91304	Homo sapiens	CORT- Human protein NOV10b.	347	93
442	gi21263092	Mus musculus	tramdorin 1	403	64
442	gi21263094	Rattus norvegicus	tramdorin 1	395	62
442	gi14571904	Rattus norvegicus	lysosomal amino acid transporter 1	358	56
443	AAU11817	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5R23V2.	877	72
443	AAU11816	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5R-3V2.	877	72
443	AAU11815	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5G-3V3.	877	72
444	gi10186503	Homo sapiens	sialic acid-specific acetylesterase	932	100

Table 2B

168

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			II		
444	gi10242345	Homo sapiens	sialic acid-specific 9-O-acetylesterase I	753	100
444	gi1628565	Mus musculus	sialic acid-specific 9-O-acetylesterase	751	81
445	gi|18087335|gb|AAL58838.1|AF390028_1	Homo sapiens	serine/threonine protein kinase kkialre-like 1	222	54
445	gi|15609263|ref|NP_216642.1|	Mycobacterium tuberculosis H37Rv	PE_PGRS	100	36
445	gi|14251041|ref|NP_116403.1|	Tupaia herpesvirus	T49	93	33
446	gi3165565	Caenorhabditis elegans	C. elegans PTR-15 protein (corresponding sequence T07H8.6)	114	23
446	gi1825729	Caenorhabditis elegans	C. elegans PTR-2 protein (corresponding sequence C32E8.8)	110	25
446	gi1255388	Caenorhabditis elegans	C. elegans PTR-1 protein (corresponding sequence C24B5.3)	83	25
447	AAB88481	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0251.	252	73
447	AAE03835	Homo sapiens	HUMA- Human gene 18 encoded secreted protein HFKHW50, SEQ ID NO: 81.	252	73
447	AAM78797	Homo sapiens	HYSE- Human protein SEQ ID NO 1459.	170	67
448	gi3130159	Takifugu rubripes	pheromone receptor	210	63
448	gi|17482335|ref|XP_064863.1|	Homo sapiens	similar to vomeronasal 2, receptor, 4; vomeronasal organ family 2, receptor, 4	448	76
448	gi|20948634|ref|XP_142573.1|	Mus musculus	similar to vomeronasal 2, receptor, 2; vomeronasal organ family 2, receptor, 2	260	79
449	gi13452508	Mus musculus	claudin 14	438	39
449	AAU77764	Homo sapiens	GETH Tumour associated antigenic target polypeptide (TAT) 155.	437	39
449	AAY99431	Homo sapiens	GETH Human PRO1571 (UNQ777) amino acid sequence SEQ ID NO:324.	437	39
450	AAM65951	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26257.	206	61
450	AAM53568	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25673.	206	61
450	AAM73342	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein	184	54

Table 2B

169

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SEQ ID NO: 33648.		
451	gi19343983	Homo sapiens	GalNAc-4-sulfotransferase 2	213	97
451	gi12711481	Homo sapiens	N-acetylgalactosamine 4-O-sulfotransferase 2 GalNAc4ST-2	187	97
451	AAM69697	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30003.	99	54
452	gi3150438	Human endogenous retrovirus K	pol-env	258	55
452	gi1469243	Human endogenous retrovirus K	pol/env	258	55
452	gi4185944	Human endogenous retrovirus K	env protein	258	55
453	gi20563599	Homo sapiens	methyl-CpG binding domain protein 3-like protein 2	982	98
453	AAU00437	Homo sapiens	COUN- Human dendritic cell membrane protein FIRE.	547	97
453	AAY91625	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 22 SEQ ID NO:298.	547	97
454	gi15590686	Homo sapiens	peptidoglycan recognition protein-l-beta precursor	1960	98
454	AAY96963	Homo sapiens	HUMA- Wound healing tissue peptidoglycan recognition protein-like protein.	1810	92
454	gi15590684	Homo sapiens	peptidoglycan recognition protein-l-alpha precursor	1223	61
455	AAE19173	Homo sapiens	INCY- Human protease, PRTS-10 protein.	1009	100
455	AAB72301	Homo sapiens	HIRO/ Human ADAMTS-9 alternative amino acid sequence.	1009	100
455	AAB72286	Homo sapiens	HIRO/ Human ADAMTS-9 amino acid sequence.	1009	100
456	ABK15497_aa 1	Homo sapiens	HOFF Human senescence associated epithelial membrane protein (SEMP1) cDNA.	150	100
456	AAZ60459_aa 1	Homo sapiens	INCY- cDNA encoding a human molecule associated with apoptosis 2 (MAPOP-2).	150	100
456	AAX19461_aa 1	Homo sapiens	UNIW Human senescence factor p23 gene.	150	100
457	gi|11191823|emb|CAC16413.1|	Streptomyces olivaceus	elloramycin glycosyltransferase	70	47
457	gi|21301888|gb|EAA14033.1|	Anopheles gambiae str. PEST	agCP8508	70	53
457	gi|17737304|ref|NP_511114.1|	Drosophila melanogaster	sevenless	70	32
458	AAM65947	Homo sapiens	MOLE- Human bone marrow	158	14

Table 2B

170

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			expressed probe encoded protein SEQ ID NO: 26253.		
458	AAM53564	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25669.	158	14
458	gi4406172	Human herpesvirus 4	latent membrane protein-1	149	36
459	AAB93188	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12140.	251	83
459	AAB92702	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11102.	251	83
459	AAM00899	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 375.	251	83
460	AAG68349	Homo sapiens	BODA- Human retinitis pigmentosa related protein 14 SEQ ID NO:2.	345	100
460	gi18175295	Homo sapiens	CRB1 isoform II precursor	345	100
460	gi18182323	Mus musculus	crumbs-like protein 1 precursor	247	71
461	AAM65406	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 37511.	277	100
461	AAM96299	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4957.	171	94
461	gi396416	Escherichia coli	similar to Neurospora crassa phosphate-repressible phosphate permease	71	37
462	gi7677068	Homo sapiens	endomembrane protein emp70 precursor isolog	73	35
463	AAB95530	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18126.	233	100
463	AAB93627	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13102.	195	80
463	gi2827162	Rattus norvegicus	rsec15	195	80
464	gi19171152	Homo sapiens	ADAMTS-19	1321	98
464	AAE10350	Homo sapiens	PFIZ Human ADAMTS-J1.4 variant protein.	205	46
464	AAE10348	Homo sapiens	PFIZ Human ADAMTS-J1.2 variant protein.	205	46
466	AAD12602_aa 1	Homo sapiens	SAGA Human protein having hydrophobic domain encoding cDNA clone HP10797.	354	100
466	AAB88353	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0079.	354	100
466	AAG81285	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:88.	354	100
467	gi|5729792|ref| NP_006577.1|	Homo sapiens	trinucleotide repeat containing 5; CAG repeat containing; expanded repeat domain, CAG/CTG 5; CAG repeat domain	67	39
467	gi|15229956|re	Arabidopsis	omega-6 fatty acid desaturase,	67	36

Table 2B

171

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	f|NP_187819.1 |	thaliana	endoplasmic reticulum (FAD2)		
467	gi|6969163|emb|CAB75301.1 |	Homo sapiens	dJ475N16.1 (CTG4A)	67	39
468	AAB38330	Homo sapiens	HUMA- Human secreted protein encoded by gene 10 clone HTEBV72.	214	97
468	gi|20341041|ref|XP_110311.1 |	Mus musculus	RIKEN cDNA 4933424G06	109	48
469	AAM95018	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3676.	488	100
469	gi13311009	Homo sapiens	NYD-SP16	488	100
469	gi1418266	Chlamydomonas eugametos	SF-assemblin	75	32
470	AAE06592	Homo sapiens	SAGA Human protein having hydrophobic domain, HP03884.	357	100
470	AAB13343	Homo sapiens	LEXI- Human cortexin-like protein.	203	59
470	ABB05043	Homo sapiens	CURA- Human NOV5a protein SEQ ID NO:22.	175	53
471	gi13938651	Mus musculus	Similar to conserved membrane protein at 44E	502	83
471	gi16768782	Drosophila melanogaster	LD03322p	443	68
471	gi14194169	Arabidopsis thaliana	At1g05960/T21E18_20	120	30
472	gi310100	Rattus norvegicus	developmentally regulated protein	536	80
472	ABB17427	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6084.	455	100
472	AAW52812	Homo sapiens	INCY- Human induced tumour protein.	227	37
473	AAI67941_aa1	Homo sapiens	FARB Human dopamine-like G protein-coupled receptor (GPCR) encoding cDNA.	1711	100
473	AAD30728_aa1	Homo sapiens	PFIZ Human G-protein coupled receptor (GPCR), PFI-007 cDNA.	1711	100
473	AAD06020_aa1	Homo sapiens	MERE Human G-protein coupled receptor, GPCR_KD5 cDNA.	1711	100
474	AAE20142	Homo sapiens	MERE Human protein containing ring finger domain, R1P4.	1593	85
474	gi13872241	Homo sapiens	bA400I20.1 (ligand of numb-protein X)	1593	85
474	gi15282065	Mus musculus	LNX2	1478	79
475	AAB08872	Homo sapiens	INCY- Amino acid sequence of a human secretory protein.	77	93
475	AAB73980	Homo sapiens	GLAX Human stargazin-like	75	29

Table 2B

172

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			protein CACNG4.		
475	AAU08723	Homo sapiens	GEMY Human clone ho1143_20 secretory protein.	75	29
477	gi10799398	Homo sapiens	kallikrein 13	1513	100
477	gi6063386	Homo sapiens	kallikrein-like protein 4 KLK-L4	1513	100
477	AAZ22639_aa 1	Homo sapiens	SMIK CASB12 derived from Expressed Sequence Tag sequences.	678	48
478	ABB08214	Homo sapiens	ZYMO Human Zsig47 protein.	704	100
478	AAU83634	Homo sapiens	GETH Human PRO protein, Seq ID No 86.	704	100
478	AAB90662	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO: 205.	704	100
479	AAO00662	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 14554.	90	72
479	gi6715140	Drosophila melanogaster	split ends	89	46
479	gi6979936	Drosophila melanogaster	split ends long isoform	89	46
480	gi17944167	Drosophila melanogaster	GH10778p	76	31
480	gi2340108	Zea mays	starch branching enzyme IIa	74	33
480	gi2764762	Amycolatopsis mediterranei	rifamycin polyketide synthase, type 1	74	32
481	ABB90747	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 226.	139	30
481	ABB50291	Homo sapiens	USSH Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.	139	30
481	AAW12843	Homo sapiens	UYMA- Pro-alpha1(III):(I) CP chimeric protein.	139	30
482	AAB65246	Homo sapiens	GETH Human PRO1100 (UNQ546) protein sequence SEQ ID NO:299.	787	100
482	AAG81355	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:228.	787	100
482	AAY66723	Homo sapiens	GETH Membrane-bound protein PRO1100.	787	100
483	AAB08216	Homo sapiens	STRD A protein related to Drosophila naked cuticle polypeptide.	247	56
483	gi16303260	Homo sapiens	Dvl-binding protein NKD1	247	56
483	gi17978537	Homo sapiens	naked protein	247	56
484	gi3452275	Pseudopleuronectes americanus	aminopeptidase N	210	28
484	gi2766187	Gallus gallus	aminopeptidase Ey	175	26
484	gi544755	Oryctolagus cuniculus	aminopeptidase N; APN	174	26
485	AAB58305	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 643.	273	100
485	gi|17562350|re	Caenorhabditi	K07C11.10.p	103	42

Table 2B

173

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	f|NP_505121.1 |	s elegans			
485	gi|17565456|ref|NP_507945.1 |	Caenorhabditis elegans	Y38H6C.3.p	79	29
486	AAB38019	Homo sapiens	HUMA- Human secreted protein encoded by gene 27 clone HPJBF63.	583	99
486	AAB38010	Homo sapiens	HUMA- Human secreted protein encoded by gene 27 clone HOUHD63.	576	98
486	gi17742574	Agrobacterium tumefaciens str. C58 (U. Washington)	monooxygenase	79	41
487	AAY91385	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 40 SEQ ID NO:106.	969	100
487	AAU75555	Homo sapiens	BIOJ Immunoglobulin superfamily member GP286a.	959	99
487	AAU83610	Homo sapiens	GETH Human PRO protein, Seq ID No 38.	959	99
488	gi15779156	Homo sapiens	Similar to RIKEN cDNA 1810073N04 gene	262	96
488	gi9971734	Galleria mellonella	heavy-chain fibroin	116	34
488	gi13880674	Mycobacterium tuberculosis CDC1551	PE_PGRS family protein	98	30
489	gi409995	Rattus sp.	mucin	167	64
489	AAM65950	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26256.	146	61
489	AAM53567	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 25672.	146	61
490	gi1841555	Homo sapiens	NG5	422	100
490	ABB90246	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2622.	119	40
490	ABB90038	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2414.	119	28
491	AAU07370	Homo sapiens	PHAA G protein-coupled receptor.	117	30
491	gi5732924	Toxocara canis	excretory/secretory mucin MUC-4	114	32
491	gi5732920	Toxocara canis	excretory/secretory mucin MUC-2	110	32
492	AAB80245	Homo sapiens	GETH Human PRO257 protein.	395	100
492	AAB70534	Homo sapiens	CURA- Human PRO4 protein sequence SEQ ID NO:8.	395	100
492	AAU12343	Homo sapiens	GETH Human PRO257 polypeptide sequence.	395	100
493	gi17861670	Drosophila melanogaster	GH20388p	159	30

Table 2B

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
493	AAM76340	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36646.	138	40
493	AAM63526	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35631.	138	40
494	AAU83220	Homo sapiens	ZYMO Novel secreted protein Z912187G1P.	1208	100
494	AAO01373	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15265.	101	51
494	AAM79091	Homo sapiens	HYSE- Human protein SEQ ID NO 1753.	88	29
495	gi1841555	Homo sapiens	NG5	80	42
495	AAB18976	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	67	40
495	gi|20823606|ref|XP_140861.1 |	Mus musculus	similar to SURF-1 protein - mouse	231	100
496	gi9885193	Homo sapiens	dJ881L22.3 (novel protein similar to a trypsin inhibitor)	1429	100
496	gi2943716	Homo sapiens	25 kDa trypsin inhibitor	840	63
496	gi13241970	Gallus gallus	SugarCrisp	839	58
497	gi306316	Herpesvirus papio	EBNA-2	163	36
497	gi4096372	Rattus norvegicus	SH3 domain binding protein	158	28
497	gi4096360	Rattus norvegicus	CR16	158	28
498	AAU81976	Homo sapiens	INCY- Human secreted protein SECP2.	373	100
498	ABB89091	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1467.	373	100
498	AAB70538	Homo sapiens	CURA- Human PRO8 protein sequence SEQ ID NO:16.	373	100
499	AAE02938	Homo sapiens	MILL- Human adenylate cyclase 25678.	262	100
499	AAB02006	Homo sapiens	TEXA Adenylyl cyclase type II-C2 C2 alpha domain.	258	98
499	gi202752	Rattus norvegicus	adenylyl cyclase type II	258	98
500	AAB93931	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13927.	1083	70
500	gi10440418	Homo sapiens	FLJ00044 protein	1083	70
500	gi16648518	Drosophila melanogaster	SD09360p	133	26
501	AAO14401	Homo sapiens	ELIL Novel human cerebellin-like protein (LP232).	493	100
501	AAE16346	Homo sapiens	CURA- Human cerebellin-like protein, POLY10.	492	98
501	ABB84924	Homo sapiens	GETH Human PRO1382 protein sequence SEQ ID NO:216.	492	98
502	gi19264106	Mus musculus	RIKEN cDNA 2810049G06 gene	1079	51
502	gi19343843	Mus musculus	Similar to RIKEN cDNA	1005	46

Table 2B

175

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			2810049G06 gene		
502	AAB93797	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13560.	956	53
503	AAG89306	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 426.	82	32
503	gi5869819	Globodera pallida	NADH-ubiquinone oxidoreductase subunit 1	78	34
503	gi15026129	Clostridium acetobutylicum	Predicted membrane protein	78	32
504	AAO14201	Homo sapiens	INCY- Human transporter and ion channel TRICH-18.	1597	99
504	AAY34120	Homo sapiens	AXYS- Human potassium channel K+Hnov4.	1597	99
504	gi16611600	Homo sapiens	voltage gated potassium channel Kv3.2a	1597	99
505	gi5902892	Streptomyces avermitilis	type I polyketide synthase AVES 2	76	26
505	AAM25582	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1097.	73	27
505	ABB11449	Homo sapiens	HYSE- Human PI3-kinase homologue, SEQ ID NO:1819.	73	27
506	gi1049106	Homo sapiens	dystonin isoform 2	76	52
506	gi904022	Mus musculus	dystonin isoform 2	72	50
507	AAB94709	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15705.	89	29
507	AAU91279	Homo sapiens	CURA- Human NOV3a protein.	88	33
507	gi6319138	Rattus norvegicus	ALG-2 interacting protein 1	84	36
508	gi2564916	Homo sapiens	cotel	231	31
509	AAE01854	Homo sapiens	HUMA- Human gene 17 encoded secreted protein fragment, SEQ ID NO:180.	371	82
509	AAE01829	Homo sapiens	HUMA- Human gene 17 encoded secreted protein HWBEM18, SEQ ID NO:150.	371	82
509	AAE01786	Homo sapiens	HUMA- Human gene 17 encoded secreted protein HWBEM18, SEQ ID NO:107.	371	82
510	ABB04347	Homo sapiens	SHAN- Human protein phosphatase 4 regulatory subunit 37.	614	95
510	gi11136904	Homo sapiens	bA109J9.1 (isoform 2 of PRO1085 protein, similar to protein serine/threonine phosphatase 4 regulatory subunit 1 (PPP4R1))	601	92
510	AAB73355	Homo sapiens	MIYA/ Human mesangial cell meg-1 protein.	312	51
511	AAE03548	Homo sapiens	FARB Human mitochondrial deformylase full length protein.	1300	100
511	gi13195254	Homo sapiens	polypeptide deformylase-like protein	1300	100
511	gi11320944	Homo sapiens	peptide deformylase-like protein	1300	100

Table 2B

176

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
512	ABK15715_aa 1	Homo sapiens	MILL- Human 21612 alcohol dehydrogenase (ADH) cDNA.	520	63
512	AAU76223	Homo sapiens	MILL- Human 21612 alcohol dehydrogenase (ADH) protein.	520	63
512	AAB84367	Homo sapiens	MILL- Amino acid sequence of human alcohol dehydrogenase 21612.	520	63
513	AAB31209	Homo sapiens	GETH Amino acid sequence of human polypeptide PRO941.	863	100
513	AAB44281	Homo sapiens	GETH Human PRO941 (UNQ478) protein sequence SEQ ID NO:264.	863	100
513	AAY41725	Homo sapiens	GETH Human PRO941 protein sequence.	863	100
514	AAB08944	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 19 SEQ ID NO:101.	206	83
514	AAB08909	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 19 SEQ ID NO:66.	159	80
514	gi15157368	Agrobacteriu m tumefaciens str. C58 (Cereon)	AGR_C_4035p	68	30
516	gi340002	Homo sapiens	thyrotropin beta subunit	739	99
516	gi7690113	Homo sapiens	thyroid-stimulating hormone beta subunit	736	98
516	AAR99419	Homo sapiens	GENZ TSH beta subunit.	723	98
517	AAB53436	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:976.	368	97
517	AAB25691	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 27 SEQ ID NO:80.	168	93
517	AAG89263	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 383.	84	33
518	AAU83188	Homo sapiens	ZYMO Novel secreted protein Z887042G3P.	1443	100
518	AAB85336	Homo sapiens	CHIR Human oaf protein sequence.	1443	100
518	AAE03851	Homo sapiens	HUMA- Human gene 8 encoded secreted protein HBIOH81, SEQ ID NO:97.	1437	99
519	AAG76189	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6953.	349	100
520	AAB88377	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0113.	379	91
520	ABB06152	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:244.	137	85
520	gi|17465349|re f|XP_069720.1 |	Homo sapiens	similar to solute carrier family 29 (nucleoside transporters), member 1	425	100

Table 2B

177

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
522	gi19263985	Homo sapiens	Similar to RIKEN cDNA 1300017E09 gene	737	99
522	gi19528309	Drosophila melanogaster	LD02310p	269	53
522	gi19577352	Aspergillus fumigatus	probable adrenoleukodystrophy protein	71	31
523	AAY54053	Homo sapiens	PHAA A variant of an angiogenesis-associated protein which binds plasminogen.	155	37
523	AAY54052	Homo sapiens	PHAA An angiogenesis-associated protein which binds plasminogen.	155	37
523	gi9887326	Homo sapiens	angiomotin	155	37
524	gi|11072097|gb|AAG26333.2|	Homo sapiens	MLL/GAS7 fusion protein	73	25
525	gi1504002	Homo sapiens	similar to a human major CRK-binding protein DOCK180.	1040	84
525	gi13195147	Mus musculus	HCH	949	77
525	AAW03515	Homo sapiens	SHKJ Human DOCK180 protein.	200	31
526	gi854065	Human herpesvirus 6	U88	305	47
526	AAB95124	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17122.	212	38
526	AAO02474	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16366.	212	47
527	AAG00214	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4295.	98	89
527	AAB58446	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 784.	98	89
527	AAY48278	Homo sapiens	META- Human prostate cancer-associated protein 64.	98	89
528	gi|15840618|ref|NP_335655.1|	Mycobacterium tuberculosis CDC1551	2,4-dienoyl-coA reductase	69	34
528	gi|15608315|ref|NP_215691.1|	Mycobacterium tuberculosis H37Rv	fadH	69	34
529	AAU83079	Homo sapiens	ZYMO Novel secreted protein Z1297G2P.	2111	100
529	AAB61421	Homo sapiens	MILL- Human TANGO 300 protein.	1861	90
529	AAB23618	Homo sapiens	ALPH- Human secreted protein SEQ ID NO: 36.	1859	90
530	gi6841194	Homo sapiens	HSPC272	418	66
530	gi3170533	Saccharomyces cerevisiae	nucleolar protein Nop5p	93	30
530	gi17862426	Drosophila melanogaster	LD27336p	90	29
531	gi|17485716|ref|XP_066655.1|	Homo sapiens	similar to G protein-coupled receptor 64; G protein-coupled receptor, epididymis-specific	67	30

Table 2B

178

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			(seven transmembrane family)		
532	gi14330383	Homo sapiens	sodium/calcium exchanger SCL8A3	190	94
532	AAM47745	Homo sapiens	MERE Human natrium(+)-calcium(2+) exchanger form 3 protein, HNCX3.	183	100
532	gi1552526	Rattus norvegicus	sodium-calcium exchanger form 3	178	92
533	gi58028	synthetic construct	suef protein	142	29
533	AAM82345	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:9938.	108	30
533	AAM72527	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 32833.	96	30
534	gi6958616	Human immunodefici ency virus type 1	envelope glycoprotein	69	25
535	gi16359163	Homo sapiens	Similar to RIKEN cDNA 2310014B08 gene	1128	100
535	gi18043464	Mus musculus	RIKEN cDNA 2310014B08 gene	843	75
535	AAB64401	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA33.	139	37
536	AAE13349	Homo sapiens	SENO- Human TSTP protein, 165-015D.	1561	91
536	AAE13348	Homo sapiens	SENO- Human TSTP protein, 165-015C.	520	36
536	AAE13350	Homo sapiens	SENO- Human TSTP protein, 165-015E.	249	28
537	AAU74622	Homo sapiens	UYCA- Oestrogen-regulated LIV-1 family protein AX078294_Hs.	974	90
537	AAU74621	Homo sapiens	UYCA- Oestrogen-regulated LIV-1 family protein Q15043_Hs.	974	90
537	AAB60496	Homo sapiens	INCY- Human cell cycle and proliferation protein CCYPR-44, SEQ ID NO:44.	974	90
538	AAG81279	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:76.	223	100
538	ABB90338	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2714.	126	47
538	AAB65159	Homo sapiens	GETH Human PRO180 (UNQ154) protein sequence SEQ ID NO:23.	126	47
539	AAB94271	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14691.	207	100
539	AAM94001	Homo sapiens	HELI- Human stomach cancer expressed polypeptide SEQ ID NO 72.	207	100

Table 2B

179

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
539	AAW78193	Homo sapiens	HUMA- Human secreted protein encoded by gene 68 clone H2CBJ08.	103	46
541	gi4574260	Haemophilus influenzae	outer membrane protein 26	70	29
541	gi19916386	Methanosarcina acetivorans str. C2A	H(+)-transporting ATP synthase, subunit gamma [Methanosarcina	69	32
541	gi|20916197|ref|XP_133065.1|	Mus musculus	RIKEN cDNA D630002J15	410	73
542	gi15559405	Homo sapiens	Similar to RIKEN cDNA 0610030G03 gene	1301	100
542	gi13543049	Mus musculus	Similar to RIKEN cDNA 0610030G03 gene	1147	87
542	gi18314524	Mus musculus	Similar to RIKEN cDNA 2010305C02 gene	272	31
543	gi14789599	Homo sapiens	Similar to RIKEN cDNA 2810403L02 gene	1839	100
543	gi11493522	Homo sapiens	PRO1512	1512	100
543	AAB58871	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 579.	1407	93
544	gi693811	Homo sapiens	VPre-B protein	788	100
544	gi2114308	Homo sapiens	VpreB	788	100
544	gi340305	Homo sapiens	VpreB protein precursor	751	99
545	gi21430832	Drosophila melanogaster	SD18306p	357	41
545	gi15823977	Streptomyces avermitilis	modular polyketide synthase	79	29
545	gi19879682	Homo sapiens	LIM homeobox transcription factor 1 alpha variant 4AB	77	31
546	AAB65258	Homo sapiens	GETH Human PRO1153 (UNQ583) protein sequence SEQ ID NO:351.	129	34
546	AAG81325	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:168.	129	34
546	AAE06576	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10764.	129	34
547	gi153366	Streptomyces cinnamonensis	methylmalonyl-CoA small subunit	76	39
547	gi18308138	Sus scrofa	CD34 antigen	74	20
547	gi|17466082|ref|XP_070192.1|	Homo sapiens	similar to olfactory receptor MOR145-1	399	90
548	gi405956	Escherichia coli	yeeE	1138	93
548	gi1736691	Escherichia coli	Exodeoxyribonuclease I (EC 3.1.11.1) (Exonuclease I) (DNA deoxyribophosphodiesterase) (DRPase).	1014	86
548	gi405954	Escherichia coli	exonuclease I	1014	86
549	gi295196	Salmonella	level of amino acid identity	699	86

Table 2B

180

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		typhimurium	between E. coli and S.typhimurium strongly suggests authentic gene		
549	gi|13541796|ref|NP_111484.1|	Thermoplasma volcanium	Predicted transporter component	137	27
550	gi17429437	Ralstonia solanacearum	PROBABLE TRANSMEMBRANE PROTEIN	251	24
550	gi|21290142|gb|EAA02287.1|	Anopheles gambiae str. PEST	ebiP1696	121	26
550	gi|15964907|ref|NP_385260.1|	Sinorhizobium meliloti	HYPOTHETICAL TRANSMEMBRANE PROTEIN	117	27
551	gi216539	Escherichia coli	BasS	825	98
551	gi536956	Escherichia coli	basS	825	98
551	gi1790551	Escherichia coli K12	sensor protein for basR	825	98
552	gi1778505	Escherichia coli	ferric enterobactin transport protein	1021	100
552	gi1786804	Escherichia coli K12	ferric enterobactin transport protein	1021	100
552	gi13360086	Escherichia coli O157:H7	ferric enterobactin transport protein	1020	99
553	gi13363896	Escherichia coli O157:H7	dipeptide transport system permease protein 2	1114	100
553	gi349227	Escherichia coli	transmembrane protein	1114	100
553	gi466681	Escherichia coli	dppC	1114	100
554	gi4063042	Cryptosporidium parvum	GP900; mucin-like glycoprotein	359	57
554	gi2827462	Cercopithecus aethiops	hepatitis A virus cellular receptor 1 long form	314	56
554	gi2827460	Cercopithecus aethiops	hepatitis A virus cellular receptor 1 short form	314	56
555	gi13959789	Homo sapiens	lung alpha/beta hydrolase protein 1	203	88
555	gi13784946	Mus musculus	alpha/beta hydrolase-1	175	77
555	gi15488726	Mus musculus	lung alpha/beta hydrolase 1	175	77
556	AAU11390	Homo sapiens	SENO- Human T2R75 (hT2R75) polypeptide.	389	98
556	gi20336511	Homo sapiens	candidate taste receptor T2RP17	389	98
556	AAU11384	Homo sapiens	SENO- Human T2R61 (hT2R61) polypeptide.	368	91
557	gi2275592	Homo sapiens	TCRBV12S1	534	100
557	gi2218039	Homo sapiens	V_segment translation product	534	100
557	gi467921	Homo sapiens	T-cell receptor beta chain V region precursor	525	100
558	gi3093754	Neurospora crassa	AR2	73	28

Table 2B

181

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
558	gi16415400	Listeria innocua	highly similar to cytochrome D ubiquinol oxidase subunit II	72	29
558	gi16412217	Listeria monocytogenes	highly similar to cytochrome D ubiquinol oxidase subunit II	72	29
559	AAD25037_aa1	Homo sapiens	GENA- Human oncostatin M (OSM) cDNA.	1306	100
559	AAE15318	Homo sapiens	GENA- Human oncostatin M (OSM) protein.	1306	100
559	AAY87820	Homo sapiens	AMGE- Human oncostatin protein.	1306	100
560	AAB49502	Homo sapiens	HUMA- Clone HYASC03.	310	98
560	gi20071228	Mus musculus	RIKEN cDNA 2810051A14 gene	151	51
560	AAG81365	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:248.	144	48
561	AAY38432	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 3.	77	47
561	gi2114473	Mus musculus	p140mDia	73	39
561	gi3638957	Homo sapiens	sco-spondin-mucin-like; similar to P98167 (PID:g1711548); details of intron/exon structure uncertain	72	32
562	gi16504300	Salmonella enterica subsp. enterica serovar Typhi	probable membrane transport protein	789	93
562	gi9948048	Pseudomonas aeruginosa	probable transporter (membrane subunit)	557	63
562	gi7227389	Neisseria meningitidis MC58	sodium/dicarboxylate symporter family protein	484	58
563	AAO14190	Homo sapiens	INCY- Human transporter and ion channel TRICH-7.	2183	91
563	gi183298	Homo sapiens	GLUT5 protein	1329	55
563	gi12804761	Homo sapiens	solute carrier family 2 (facilitated glucose transporter), member 5	1329	55
564	AAE14571	Homo sapiens	EXEL- Human rhomboid related protein, RRP3.	576	100
564	ABB75690	Homo sapiens	SHAN- Human rhombus related protein 48-35.53.	576	100
564	gi19171162	Homo sapiens	ventrhoid transmembrane protein	576	100
565	AAD17516_aa1	Homo sapiens	SENO- Human taste receptor, hT1R1 cDNA coding sequence.	968	100
565	ABB77319	Homo sapiens	INCY- Human G-protein coupled receptor SEQ ID NO 3.	968	100
565	AAE10372	Homo sapiens	SENO- Human taste receptor, hT1R1 protein.	968	100
566	gi20147226	Arabidopsis thaliana	At2g44720/F16B22.21	101	38
566	AAO13099	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26991.	100	40

Table 2B

182

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
566	gi1762434	Sus scrofa	nitric oxide synthase	98	29
567	AAO08397	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22289.	231	93
567	AAM41207	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6138.	161	73
567	AAM39421	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2566.	161	73
568	AAU83199	Homo sapiens	ZYMO Novel secreted protein Z891639G1P.	900	100
568	AAB95726	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18602.	492	75
568	AAB95109	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17089.	492	75
569	AAU83607	Homo sapiens	GETH Human PRO protein, Seq ID No 32.	311	100
569	gi17976983	Xenopus laevis	P8F7	196	41
569	gi10176740	Arabidopsis thaliana	RING zinc finger protein-like	81	33
570	AAD20624_aa 1	Homo sapiens	HUMA- Human ovarian cancer antigen-encoding gene 7 cDNA clone HMAMI21.	437	89
570	AAB87396	Homo sapiens	HUMA- Human gene 8 encoded secreted protein HMAM121, SEQ ID NO:137.	437	89
570	AAB85784	Homo sapiens	INCY- Human kinase PKIN-3.	437	89
571	AAM86710	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:14303.	387	97
571	gi|19527326|ref|NP_598857.1 |	Mus musculus	expressed sequence AW049604	161	96
572	gi1491621	Bovine herpesvirus 1	UL36	97	36
572	gi2653311	Bovine herpesvirus type 1.1	very large virion protein (tegument)	97	36
572	ABB11413	Homo sapiens	HYSE- Human extensin homologue, SEQ ID NO:1783.	89	28
573	gi15292437	Drosophila melanogaster	LP10272p	131	39
573	AAY87336	Homo sapiens	INCY- Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	69	34
573	gi4877582	Homo sapiens	lipoma HMGIC fusion partner	69	34
574	AAM92575	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 1924.	350	98
574	gi15426530	Homo sapiens	similar to RIKEN cDNA 1810006A16 gene	212	29
574	gi15080253	Homo sapiens	Similar to RIKEN cDNA 1810006A16 gene	212	29
575	AAU83712	Homo sapiens	GETH Human PRO protein, Seq ID No 242.	724	91

Table 2B

183

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
575	gi16359053	Homo sapiens	Similar to RIKEN cDNA 2010309H15 gene	724	91
575	AAB19403	Homo sapiens	CHIR Amino acid sequence of a human secreted protein.	712	89
576	gi12718841	Mus musculus	Skullin	301	38
576	gi4191356	Mus musculus	claudin-6	299	38
576	gi13543081	Mus musculus	claudin 6	299	38
577	gi801882	Vibrio alginolyticus	FkuB	76	31
578	AAO14197	Homo sapiens	INCY- Human transporter and ion channel TRICH-14.	135	44
578	AAU91185	Homo sapiens	MILL- Human HEAT-2 polypeptide.	135	44
578	gi19527485	Drosophila melanogaster	LD19039p	119	41
579	gi18044066	Mus musculus	RIKEN cDNA 5033406L14 gene	342	84
579	gi6682873	Homo sapiens	reduced expression in cancer	200	90
579	gi7230612	Rattus norvegicus	small rec	197	87
580	AAM25339	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:854.	351	100
580	ABB89642	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2018.	291	84
580	gi2204110	Bos taurus	adenylyl cyclase type VII	233	69
581	AAB24476	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 40 SEQ ID NO:101.	238	69
581	AAM82470	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:10063.	122	71
581	gi18461301	Oryza sativa (japonica cultivar-group)	similar to 26S proteasome subunit4	86	24
582	AAE14571	Homo sapiens	EXEL- Human rhomboid related protein, RRP3.	341	100
582	ABB75690	Homo sapiens	SHAN- Human rhombus related protein 48-35.53.	341	100
582	gi19171162	Homo sapiens	ventrhoid transmembrane protein	341	100
583	AAU18887	Homo sapiens	HUMA- Novel prostate gland antigen, Seq ID No 186.	348	95
583	AAM96039	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4697.	348	95
583	gi16648412	Drosophila melanogaster	LD44720p	160	33
584	gi5305335	Mycobacterium tuberculosis	proline-rich mucin homolog	132	29
584	gi2429362	Santalum album	proline rich protein	126	28
584	gi12018147	Chlamydomonas reinhardtii	vegetative cell wall protein gp1	124	30

Table 2B

184

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
585	gi3165565	Caenorhabditis elegans	C. elegans PTR-15 protein (corresponding sequence T07H8.6)	89	29
585	gi1825729	Caenorhabditis elegans	C. elegans PTR-2 protein (corresponding sequence C32E8.8)	79	20
585	gi15824031	Streptomyces avermitilis	modular polyketide synthase	75	30
586	gi8919836	Blumeria graminis f. sp. hordei	GTPase activating protein	76	23
586	gi21064465	Drosophila melanogaster	RE36839p	75	41
586	gi2558537	Fossombronia pusilla	NADH dehydrogenase subunit 5	74	26
587	AAD02051_aa1	Homo sapiens	LEXI- Human ion channel protein (ICP) cDNA.	1195	99
587	AAE17448	Homo sapiens	MILL- Human sodium ion channel family protein, 56201.	1195	99
587	AAY71949	Homo sapiens	LEXI- Human alternative ion channel protein (ICP).	1195	99
588	gi478889	Rana catesbeiana	transcription factor RcC/EPB-1	78	33
588	gi15912317	Arabidopsis thaliana	AT4g00090/F6N15_8	75	30
588	gi|20853571|ref|XP_121991.1|	Mus musculus	similar to human immunodeficiency virus type I enhancer binding protein 2; human immunodeficiency virus type I enhancer-binding protein 2	74	34
589	AAO01188	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15080.	248	86
589	AAY73334	Homo sapiens	INCY- HTRM clone 1805061 protein sequence.	79	35
589	gi11345434	Thermus thermophilus	competence factor ComEA	78	35
590	ABB84982	Homo sapiens	GETH Human PRO5730 protein sequence SEQ ID NO:332.	603	60
590	AAM39990	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3135.	603	60
590	AAM38999	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2144.	603	60
591	AAB73674	Homo sapiens	INCY- Human oxidoreductase protein ORP-7.	193	77
592	gi36853	Homo sapiens	T-cell receptor alpha-chain (413 is 2nd base in codon)	585	100
592	gi2358022	Homo sapiens	TCRAV2S1	585	100
592	ABB11158	Homo sapiens	HYSE- Human TCR alpha chain homologue, SEQ ID NO:1528.	576	99
593	ABB90299	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2675.	121	38
593	AAM41275	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6206.	121	38

Table 2B

185

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
593	AAM39489	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2634.	121	38
594	ABB06606	Homo sapiens	CURA- G protein-coupled receptor GPCR4c protein SEQ ID NO:22.	1583	100
594	ABB06605	Homo sapiens	CURA- G protein-coupled receptor GPCR4a protein SEQ ID NO:20.	1583	100
594	ABB06604	Homo sapiens	CURA- G protein-coupled receptor GPCR4a protein SEQ ID NO:18.	1583	100
595	gi6539444	Prunus avium	S6-RNase	77	40
595	gi9957252	Prunus dulcis	Sg-RNase	77	40
595	gi9081843	Prunus dulcis	self-incompatibility associated ribonuclease	77	40
596	AAF90612_aa 1	Homo sapiens	ZYMO Human secretin-like receptor Zgpr1 cDNA.	581	100
596	AAE15635	Homo sapiens	INCY- Human G-protein coupled receptor-5 (GCREC-5) protein.	581	100
596	AAB66272	Homo sapiens	MILL- Human TANGO 378 SEQ ID NO: 29.	581	100
597	ABB84906	Homo sapiens	GETH Human PRO1287 protein sequence SEQ ID NO:180.	785	98
597	AAB65273	Homo sapiens	GETH Human PRO1287 (UNQ656) protein sequence SEQ ID NO:381.	785	98
597	AAB87561	Homo sapiens	GETH Human PRO1287.	785	98
598	gi17426446	Homo sapiens	bA351K23.5 (novel protein)	1630	100
598	ABL53627_aa 1	Homo sapiens	GENO- Breast protein-eukaryotic conserved gene 1 (BSTP-ECG1) cDNA.	909	48
598	ABB75677	Homo sapiens	GENO- Breast protein-eukaryotic conserved gene 1 (BSTP-ECG1) protein.	909	48
599	gi15012190	Homo sapiens	Similar to RIKEN cDNA 2610036L13 gene	967	94
599	AAG89274	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 394.	420	98
599	gi20987202	Mus musculus	RIKEN cDNA 2610036L13 gene	380	64
600	gi7717312	Homo sapiens	TGF-beta-activated kinase like	422	97
600	AAB18666	Homo sapiens	INCY- A human regulator of intracellular phosphorylation.	115	92
600	gi11342496	Bacteriophage phi-Ea1h	holin	75	26
601	AAG02869	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6950.	253	98
601	AAB10262	Homo sapiens	GEMY Human fetal brain protein fragment BF290_1i.	253	98
601	AAB59017	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 725.	253	98
602	gi204144	Rattus	profilaggrin	82	25

Table 2B

186

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
		norvegicus			
602	gi7682468	Bos taurus	submaxillary mucin	82	26
602	gi14578315	Plasmodium vivax	PV1H14175_P	80	26
603	gi1234787	Xenopus laevis	up-regulated by thyroid hormone in tadpoles; expressed specifically in the tail and only at metamorphosis; membrane bound or extracellular protein; C-terminal basic region	1190	55
603	AAU12201	Homo sapiens	GETH Human PRO1779 polypeptide sequence.	1181	57
603	AAB94773	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15860.	771	63
605	ABB11373	Homo sapiens	HYSE- Human olfactory receptor homologue, SEQ ID NO:1743.	482	100
605	gi18479402	Mus musculus	olfactory receptor MOR160-1	384	78
605	gi18480922	Mus musculus	olfactory receptor MOR160-4	353	74
606	gi21112746	Xanthomonas campestris pv. campestris str. ATCC 33913	C-type cytochrome biogenesis membrane protein	75	29
606	gi15128587	Inversidens japanensis	NADH dehydrogenase complex 1	72	32
606	gi|13385822|ref|NP_080601.1|	Mus musculus	RIKEN cDNA 1810059G22	727	83
607	gi13507259	Homo sapiens	amnionless	1623	75
607	AAB65237	Homo sapiens	GETH Human PRO1028 (UNQ513) protein sequence SEQ ID NO:281.	1167	99
607	AAY66714	Homo sapiens	GETH Membrane-bound protein PRO1028.	1167	99
608	AAY76219	Homo sapiens	HUMA- Human secreted protein encoded by gene 96.	336	94
608	AAY30164	Homo sapiens	ASTR- Human dorsal root receptor 6 hDRR6.	114	34
608	gi19338918	Homo sapiens	G protein-coupled receptor SNSR6	114	34
609	AAU74538	Homo sapiens	FARB Human P2Y purinoceptor 8-like G protein-coupled receptor related protein.	113	34
609	gi1771972	Xenopus laevis	P2Y8 nucleotide receptor	113	34
609	AAS19414_aa1	Homo sapiens	CURA- Human cDNA encoding novel G protein-coupled receptor, GPCR9.	102	37
610	gi15292437	Drosophila melanogaster	LP10272p	245	38
610	AAY87336	Homo sapiens	INCY- Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	101	25
610	gi4877582	Homo sapiens	lipoma HMGIC fusion partner	101	25

Table 2B

187

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
611	AAY27721	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 29.	1114	98
611	AAB87068	Homo sapiens	MILL- Human secreted protein TANGO 365, SEQ ID NO:46.	621	99
611	AAB87148	Homo sapiens	MILL- Human secreted protein TANGO 365 T20S variant, SEQ ID NO:165.	617	98
613	AAM74132	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 34438.	267	100
613	AAM61375	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 33480.	267	100
613	AAB53312	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:852.	267	100
614	AAU27619	Homo sapiens	ZYMO Human protein AFP583515.	656	89
614	AAY91598	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 8 SEQ ID NO:271.	656	89
614	AAY91458	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 8 SEQ ID NO:131.	656	89
615	gi2065210	Mus musculus	Pro-Pol-dUTPase polyprotein	1026	82
615	gi|3860513|emb|CAA13574.1|	Mus famulus	reverse transcriptase	482	84
615	gi|4379237|emb|CAA13572.1|	Mus musculus	reverse transcriptase	477	83
616	AAU25709	Homo sapiens	PHAA G protein-coupled receptor, nGPCR-2123.	136	52
617	gi13422363	Caulobacter crescentus CB15	sensor histidine kinase DivJ	76	29
617	gi699512	Mus musculus	cyclin F	75	40
617	gi7272187	Caulobacter crescentus	histidine protein kinase	74	29
618	gi15718476	Homo sapiens	Fanconi anemia complementation group D2 protein	657	90
618	gi13324523	Homo sapiens	Fanconi anemia complementation group D2 protein, isoform 2	657	90
618	gi13324522	Homo sapiens	Fanconi anemia complementation group D2 protein, isoform 1	657	90
619	AAU25447	Homo sapiens	INCY- Human mddt protein from clone LG:1083142.1:2000MAY19.	394	96
619	AAU16295	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1248.	329	95
619	AAM79834	Homo sapiens	HYSE- Human protein SEQ ID	267	65

Table 2B

188

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			NO 3480.		
620	AAB47106	Homo sapiens	ZYMO Second splice variant of MAPP.	309	51
620	gi18147612	Homo sapiens	metalloprotease disintegrin	309	51
620	gi13157560	Homo sapiens	dJ964F7.1 (novel disintegrin and reprolysin metalloproteinase family protein)	309	51
621	gi18606367	Mus musculus	RIKEN cDNA 4930570C03 gene	715	92
621	AAB90649	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO: 192.	562	97
621	AAB90565	Homo sapiens	HUMA- Human secreted protein, SEQ ID NO: 103.	472	100
622	AAY87335	Homo sapiens	INCY- Human signal peptide containing protein HSPP-112 SEQ ID NO:112.	623	99
622	gi15149556	Drosophila melanogaster	junctophilin	90	26
622	gi21428978	Drosophila melanogaster	GH28348p	90	26
623	ABB89722	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2098.	230	100
623	AAY87250	Homo sapiens	INCY- Human signal peptide containing protein HSPP-27 SEQ ID NO:27.	230	100
623	AAY92710	Homo sapiens	ZYMO Human membrane-associated protein Zsig24.	230	100
624	gi10441465	Homo sapiens	actin filament associated protein	274	90
624	gi17644147	Rattus norvegicus	actin filament associated protein	237	77
624	gi13129529	Gallus gallus	neural actin filament protein	201	71
625	gi15193279	Oncorhynchus mykiss	TNF decoy receptor	70	35
626	ABB11417	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1787.	443	98
626	AAG81257	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:32.	148	68
626	AAB12121	Homo sapiens	PROT- Hydrophobic domain protein from clone HP02962 isolated from KB cells.	148	68
627	AAM74424	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 34730.	151	73
627	AAM61632	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 33737.	151	73
627	gi13310191	multiple sclerosis associated retrovirus element	recombinant envelope protein	121	35
628	gi17390957	Mus musculus	Similar to RIKEN cDNA 2010001E11 gene	122	34
628	gi|20858407|re	Mus musculus	RIKEN cDNA 2010001E11	129	35

Table 2B

189

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
	f|XP_125582.1 |				
629	AAE03765	Homo sapiens	HUMA- Human gene 2 encoded secreted protein HCE3C63, SEQ ID NO:35.	194	58
629	gi20988899	Mus musculus	similar to deleted in bladder cancer chromosome region candidate 1	190	56
629	AAV83819_aa 1	Homo sapiens	CURI- Tumour suppressor gene IB3089A (also known as DBCCR1).	148	56
630	gi14276188	Desulfosarcin a variabilis	dissimilatory sulfite reductase alpha subunit	75	48
630	gi14669513	uncultured bacterium	dissimilatory sulfite reductase alpha subunit	74	47
630	gi14090308	Desulfovibrio cuneatus	sulfite reductase alpha subunit	73	48
631	AAM74983	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35289.	269	91
631	AAM62179	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 34284.	269	91
631	AAU20596	Homo sapiens	HUMA- Human secreted protein, Seq ID No 588.	137	50
632	AAE10339	Homo sapiens	ENGE- Human cholecystokinin (CCK).	356	100
632	AAB24381	Homo sapiens	ALLR Human procholecystokinin amino acid sequence SEQ ID NO:1.	356	100
632	gi179996	Homo sapiens	cholecystokinin	356	100
633	AAM69871	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30177.	228	100
633	AAM57476	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29581.	228	100
633	AAM68493	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28799.	223	85
634	gi14456239	Homo sapiens	bA74P14.2 (novel protein)	1684	88
634	gi4097231	Ureaplasma urealyticum	multiple banded antigen	395	23
634	gi600118	Zea mays	extensin-like protein	324	35
635	AAE05188	Homo sapiens	INCY- Human drug metabolising enzyme (DME-19) protein.	171	51
635	AAB12140	Homo sapiens	PROT- Hydrophobic domain protein isolated from WERI-RB cells.	171	51
635	ABB11624	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1994.	144	52
636	gi18676590	Homo sapiens	FLJ00193 protein	1339	98
636	AAB66267	Homo sapiens	MILL- Human TANGO 272	1326	97

Table 2B

190

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			SEQ ID NO: 14.		
636	gi17386053	Mus musculus	Jedi protein	987	79
637	gi7542324	Erwinia amylovora	potential ORFB-specific chaperone	72	27
638	AAO14184	Homo sapiens	INCY- Human transporter and ion channel TRICH-1.	1430	100
638	gi13926111	Homo sapiens	2P domain potassium channel Talk-2	1430	100
638	AAE01027	Homo sapiens	MILL- Human TWIK-3 protein from clone Athua133f10.	1426	99
639	gi2754696	Gallus gallus	high molecular mass nuclear antigen	93	27
639	gi437055	Macaca mulatta	mucin	92	29
639	gi6715140	Drosophila melanogaster	split ends	91	30
641	gi3127176	Homo sapiens	sulfonylurea receptor 2B	713	98
641	gi3127175	Homo sapiens	sulfonylurea receptor 2A	713	98
641	gi15778680	Oryctolagus cuniculus	sulphonylurea receptor 2B	678	93
642	AAB24035	Homo sapiens	GETH Human PRO4397 protein sequence SEQ ID NO:42.	1894	100
642	gi17225044	Mus musculus	beta-1,3-galactosyltransferase-related protein	1487	82
642	AAY93951	Homo sapiens	HUMA- Amino acid sequence of a Brainiac-5 polypeptide.	1241	100
643	ABB84966	Homo sapiens	GETH Human PRO4371 protein sequence SEQ ID NO:300.	758	95
643	ABB89971	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2347.	758	95
643	AAU12250	Homo sapiens	GETH Human PRO4371 polypeptide sequence.	758	95
644	AAU10355	Homo sapiens	LEEH/ Hhuman apolipoprotein C-IV (APOC4).	693	100
644	gi975893	Homo sapiens	apoC-IV	693	100
644	gi18088771	Homo sapiens	apolipoprotein C-IV	693	100
645	AAM25873	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1388.	110	80
645	AAY57878	Homo sapiens	INCY- Human transmembrane protein HTMPN-2.	101	86
646	AAG93311	Homo sapiens	NISC- Human protein HP10562.	488	100
646	AAG67820	Homo sapiens	SHAN- Human leucine zipper protein 43.	488	100
646	AAY64650	Homo sapiens	GEST Human luman homology protein.	488	100
647	gi11935177	Mus musculus	heparin/heparan sulfate:glucuronic acid C5 epimerase	1001	94
647	gi13442978	Mus musculus	D-glucuronyl C5-epimerase	1001	94
647	gi13654639	Bos taurus	D-glucuronyl C5 epimerase	969	92
648	AAG00122	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4203.	102	100
648	AAW70542	Homo sapiens	TORA Integrin alpha-2 chain.	102	100
648	gi33907	Homo sapiens	integrin alpha-2 preprotein (AA	102	100

Table 2B

191

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			-29 to 1152)		
649	gi21107282	Xanthomonas axonopodis pv. citri str. 306	TonB-dependent receptor	70	31
650	ABB90225	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2601.	683	100
650	AAB12150	Homo sapiens	PROT- Hydrophobic domain protein isolated from HT-1080 cells.	683	100
650	ABB06157	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:249.	675	98
651	AAV03875_aa 1	Homo sapiens	BETH- HTm4 gene.	173	100
651	AAW41056	Homo sapiens	BETH- HTm4 protein.	173	100
651	gi561639	Homo sapiens	IgE receptor beta subunit	173	100
652	gi21483462	Drosophila melanogaster	LD44686p	140	37
652	AAB67576	Homo sapiens	INCY- Amino acid sequence of a human hydrolytic enzyme HYENZ8.	104	43
652	AAM40456	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5387.	104	43
653	gi7209315	Homo sapiens	FLJ00007 protein	1375	85
653	AAM90874	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18467.	591	99
653	AAY99428	Homo sapiens	GETH Human PRO1431 (UNQ737) amino acid sequence SEQ ID NO:315.	430	93
654	gi297172	Rattus rattus	ribosomal protein S7	432	93
654	gi551251	Homo sapiens	ribosomal protein S7	432	93
654	gi2811284	Mus musculus	ribosomal protein S7	432	93
655	AAB68888	Homo sapiens	INCY- Human RECAP polypeptide, SEQ ID NO: 18.	273	71
655	AAU12284	Homo sapiens	GETH Human PRO5993 polypeptide sequence.	273	71
655	AAB82854	Homo sapiens	FARB Human P2Y-like GPCR polypeptide.	164	75
656	gi4096055	Homo sapiens	R28379_3	136	100
656	gi9947429	Pseudomonas aeruginosa	heme exporter protein CcmB	79	31
656	gi2984101	Aquifex aeolicus	nodulation competitiveness protein NfeD	77	28
657	AAU16396	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1349.	97	41
657	AAB95861	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18926.	96	42
657	gi6690339	Mus musculus	hematopoietic zinc finger protein	94	40
658	AAG67525	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	1850	100
658	ABB90207	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2583.	558	38
658	AAB69185	Homo sapiens	SREN- Human hISLR-iso	558	38

Table 2B

192

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			protein SEQ ID NO:7.		
659	AAH77293_aa 1	Homo sapiens	MILL- Human ion channel protein IC32391 cDNA coding region.	505	98
659	AAE13278	Homo sapiens	INCY- Human transporters and ion channels (TRICH)-5.	505	98
659	AAG77969	Homo sapiens	MILL- Human ion channel protein IC32391.	505	98
660	AAU11356	Homo sapiens	SCHE Human DNAX cytokine receptor subunit 9 (DCRS9) polypeptide.	1120	89
660	AAU83601	Homo sapiens	GETH Human PRO protein, Seq ID No 20.	1116	97
660	AAU04957	Homo sapiens	GETH Human Interleukin 17 receptor, IL-17RH3.	1116	97
661	gi1504002	Homo sapiens	similar to a human major CRK-binding protein DOCK180.	548	78
661	gi13195147	Mus musculus	HCH	514	73
661	gi1339910	Homo sapiens	DOCK180 protein	436	60
662	AAY27669	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 103.	255	100
662	gi|19527364|ref|NP_598894.1 |	Mus musculus	expressed sequence AI195350; cDNA sequence, clone 2-37	257	83
663	gi9971784	Bovine ephemeral fever virus	protein L	73	27
663	gi155287	Vibrio cholerae	disulfide isomerase	70	28
663	gi|10086573|ref|NP_065409.1 |	Bovine ephemeral fever virus	protein L	73	27
664	gi6822060	Arabidopsis thaliana	peptide transport-like protein	86	31
664	gi20147231	Arabidopsis thaliana	At1g68570/F24J5_7	74	36
664	gi20453068	Arabidopsis thaliana	At2g40460/T2P4.19	72	31
665	AAE17537	Homo sapiens	INCY- Human protein modification and maintenance molecule-6 (PMMM-6).	2583	100
665	gi21388771	Homo sapiens	kringle-containing protein	2220	100
665	gi21388540	Mus musculus	Kremen2 protein	2140	85
666	ABB07527	Homo sapiens	INCY- Human drug metabolizing enzyme (DME) (ID: 5643401CD1).	659	87
666	ABB07515	Homo sapiens	INCY- Human drug metabolizing enzyme (DME) (ID: 8097779CD1).	565	99
666	gi13161409	Mus musculus	family 4 cytochrome P450	437	73
667	AAB08862	Homo sapiens	INCY- Amino acid sequence of a human secretory protein.	958	100
667	AAB12163	Homo sapiens	PROT- Hydrophobic domain protein from clone HP10671	953	99

Table 2B

193

SEQ ID NO:	Accession No.	Species	Description	Score	% Identity
			isolated from Thymus cells.		
667	AAE10183	Homo sapiens	HYSE- Human bone marrow derived protein, SEQ ID NO: 27.	268	91
668	gi15292437	Drosophila melanogaster	LP10272p	361	31
668	AAY87336	Homo sapiens	INCY- Human signal peptide containing protein HSPP-113 SEQ ID NO:113.	181	28
668	gi4877582	Homo sapiens	lipoma HMGIC fusion partner	181	28
669	gi3598974	Rattus norvegicus	protein tyrosine phosphatase TD14	103	38
669	ABB03068	Homo sapiens	HUMA- Human expressed polypeptide SEQ ID NO 41.	83	35
669	AAB29664	Homo sapiens	KYOW Human tyrosine phosphatase HD-PTP cKAL11 fragment.	83	35
670	gi18375957	Neurospora crassa	related to 60s ribosomal protein L2 (mitochondrial)	74	32
670	gi|20540703|ref|XP_046834.6|	Homo sapiens	serologically defined colon cancer antigen 43	83	32
670	gi|18375957|emb|CAD21256.1|	Neurospora crassa	related to 60s ribosomal protein L2 (mitochondrial)	74	32
671	gi12656590	Danio rerio	P2x purinoceptor subunit 4	72	40
671	gi2995988	Callithrix jacchus	NADH dehydrogenase subunit 4	70	28
671	gi2995982	Callithrix pygmaea	NADH dehydrogenase subunit 4	70	28
672	gi1196439	Homo sapiens	latent transforming growth factor-binding protein	291	98
672	gi207286	Rattus norvegicus	TGF-beta masking protein large subunit	226	77
672	gi3493176	Mus musculus	latent TGF beta binding protein	217	73

Table 3
194

SEQ ID NO:	Database entry ID	Description	*Results
339	PR00709	AVIDIN SIGNATURE	PR00709A 4.60 1.170e-09 16-34
340	BL01253	Type I fibronectin domain proteins.	BL01253F 14.35 5.050e-14 78-116
346	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.339e-12 4-29
346	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 3.769e-10 6-29
354	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 6.362e-29 129-167
354	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354N 13.17 5.661e-12 196-240 DM01354M 12.50 1.000e-11 171-200
356	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463B 17.50 3.314e-13 135-156 PR00463A 11.40 8.568e-10 111-130
362	BL00211	ABC transporters family proteins.	BL00211B 13.37 2.286e-13 222-253 BL00211A 12.23 9.550e-09 160-171
378	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.780e-09 78-92
384	PF00624	Flocculin repeat proteins.	PF00624J 6.21 7.070e-09 40-94 PF00624F 11.04 9.056e-09 68-103
386	BL01303	BCCT family of transporters proteins.	BL01303A 14.33 5.629e-31 89-121 BL01303B 10.14 2.250e-18 142-160
387	PR00075	FATTY ACID DESATURASE FAMILY 1 SIGNATURE	PR00075A 16.97 9.565e-09 9-29
391	BL00538	Bacterial chemotaxis sensory transducers proteins.	BL00538C 10.61 1.000e-40 152-190 BL00538A 23.61 3.647e-39 96-143
391	PR00260	BACTERIAL CHEMOTAXIS SENSORY TRANSDUCER SIGNATURE	PR00260A 13.20 3.172e-24 5-30 PR00260D 9.90 4.418e-19 143-172 PR00260C 10.26 3.302e-11 69-89 PR00260B 8.90 2.220e-10 60-75
394	BL00077	Heme-copper oxidase catalytic subunit, copper B binding regio.	BL00077C 18.98 9.697e-09 9-59
400	PR00550	HYPERGLYCEMIC HORMONE SIGNATURE	PR00550C 11.31 9.426e-10 29-39
402	DM01283	A-BINDING PROTEIN CHLOROPHYLL.	DM01283A 14.91 9.600e-10 35-70
402	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 6.056e-11 57-71 PR00456E 3.06 2.367e-09 62-76 PR00456E 3.06 3.278e-09 56-70 PR00456E 3.06 4.646e-09 49-63
402	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 4.875e-10 59-73 PR00833H 2.30 2.154e-09 38-52 PR00833H 2.30 3.538e-09 88-102 PR00833H 2.30 5.615e-09 92-106 PR00833H 2.30 7.692e-09 97-111
402	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306B 5.57 9.000e-09 90-100
402	PF00624	Flocculin repeat proteins.	PF00624F 11.04 9.347e-09 85-120
402	PD01364	MUCIN GLYCOPROTEIN PRECURSOR MEM.	PD01364B 13.94 9.526e-09 109-124
402	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308A 5.90 2.694e-09 91-105 PR00308A 5.90 9.788e-09 62-76

Table 3
195

SEQ ID NO:	Database entry ID	Description	*Results
402	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191D 13.94 9.922e-09 86-124
404	BL00649	G-protein coupled receptors family 2 proteins.	BL00649B 20.68 5.061e-11 23-68 BL00649C 17.82 4.955e-10 82-107
404	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 5.435e-09 84-107 PR00249A 15.88 6.642e-09 18-42
406	BL00312	Glycophorin A proteins.	BL00312B 9.22 9.911e-09 2-30
408	PR00957	GENE 66 (IR5) PROTEIN SIGNATURE	PR00957A 7.65 3.473e-09 158-175
409	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479A 19.86 1.220e-10 59-81
409	PD02269	CYTIDINE DEAMINASE HYDROLASE ZINC AMINOHY.	PD02269C 16.36 9.735e-10 70-82
410	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 7.698e-13 116-135 PR00007D 9.64 9.654e-11 193-203 PR00007A 19.33 2.552e-10 89-115 PR00007C 15.60 3.656e-10 163-184
410	BL01113	C1q domain proteins.	BL01113B 18.26 1.563e-20 95-130 BL01113D 7.47 9.308e-12 195-204 BL01113C 13.18 4.750e-10 163-182
412	PR00925	NONHISTONE CHROMOSOMAL PROTEIN HMG17 FAMILY SIGNATURE	PR00925B 3.73 5.982e-10 78-90
414	BL00019	Actinin-type actin-binding domain proteins.	BL00019D 15.33 3.948e-14 41-70
426	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 7.000e-14 67-106
426	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 6.143e-12 36-57 PR00245B 10.38 1.675e-11 154-168
426	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 1.000e-09 81-103
435	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011B 13.08 5.576e-13 76-94 PR00011D 14.03 6.943e-13 76-94 PR00011B 13.08 9.542e-13 33-51 PR00011D 14.03 3.211e-12 33-51 PR00011A 14.06 6.516e-12 33-51 PR00011A 14.06 8.548e-12 76-94 PR00011D 14.03 3.213e-11 162-180 PR00011B 13.08 2.174e-10 162-180 PR00011D 14.03 2.523e-10 119-137 PR00011B 13.08 2.356e-09 119-137 PR00011B 13.08 5.685e-09 205-223 PR00011A 14.06 6.425e-09 119-137 PR00011A 14.06 6.671e-09 162-180 PR00011D 14.03 9.870e-09 205-223
441	PR00251	BACTERIAL OPSIN SIGNATURE	PR00251G 16.33 4.000e-09 176-194
441	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308A 5.90 6.188e-09 51-65

Table 3
196

SEQ ID NO:	Database entry ID	Description	*Results
447	BL01144	Ribosomal protein L31e proteins.	BL01144 25.07 6.684e-17 83-134
448	BL00979	G-protein coupled receptors family 3 proteins.	BL00979M 14.39 6.532e-11 30-80
448	PR00248	METABOTROPIC GLUTAMATE GPCR SIGNATURE	PR00248F 14.25 1.923e-10 56-78
453	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 6.073e-13 21-46
453	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 9.129e-11 23-46
458	BL00242	Integrins alpha chain proteins.	BL00242E 9.03 8.154e-09 82-110
458	PR00336	LYSOSOME-ASSOCIATED MEMBRANE GLYCOPROTEIN SIGNATURE	PR00336D 9.96 1.000e-08 133-155
464	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 8.071e-10 122-154
464	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 2.607e-09 142-154
464	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.714e-11 140-154 PR00049D 0.00 4.356e-09 135-149
468	PR00806	VINCULIN SIGNATURE	PR00806C 11.07 8.839e-09 13-30
473	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 9.129e-15 71-110 BL00237C 13.19 1.346e-13 218-244 BL00237D 11.23 9.308e-11 271-287
473	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 3.520e-13 223-247 PR00237C 15.69 2.200e-11 85-107 PR00237E 13.03 2.588e-09 166-189 PR00237G 19.63 3.093e-09 261-287
477	BL00495	Apple domain proteins.	BL00495N 11.04 8.239e-14 204-238 BL00495O 13.75 9.000e-14 236-264
477	BL00134	Serine proteases, trypsin family, histidine proteins.	BL00134B 15.99 4.176e-22 212-235 BL00134A 11.96 7.158e-19 61-77 BL00134C 13.45 6.850e-13 245-258
477	PR00722	CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1) SIGNATURE	PR00722A 12.27 5.737e-17 62-77 PR00722C 10.87 4.600e-15 211-223 PR00722B 12.51 4.375e-12 120-134
477	BL01253	Type I fibronectin domain proteins.	BL01253G 11.34 8.352e-14 211-224 BL01253D 4.84 7.207e-13 61-74 BL01253H 13.15 7.124e-12 227-261
477	BL00021	Kringle domain proteins.	BL00021B 13.33 2.565e-17 61-78 BL00021D 24.56 1.110e-10 217-258
477	PR00839	V8 SERINE PROTEASE FAMILY SIGNATURE	PR00839B 11.20 3.955e-10 61-78
477	BL00672	Serine proteases, V8 family, histidine proteins.	BL00672A 9.79 1.120e-09 61-76
496	PR00838	VENOM ALLERGEN 5 SIGNATURE	PR00838G 16.07 9.760e-12 165-184 PR00838D 8.73 1.563e-10 87-105
496	BL01009	Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins.	BL01009D 14.19 2.976e-17 167-187 BL01009A 13.75 3.057e-11 87-104 BL01009E 13.50 2.125e-10 201-216
496	PR00837	ALLERGEN V5/TPX-1 FAMILY SIGNATURE	PR00837C 17.21 1.000e-16 166-182 PR00837A 14.77 3.919e-13 87-105

Table 3
197

SEQ ID NO:	Database entry ID	Description	*Results
			PR00837D 11.12 9.514e-10 202-215
497	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 7.344e-13 205-219 PR00049D 0.00 9.262e-13 206-220 PR00049D 0.00 4.000e-12 207-221 PR00049D 0.00 4.000e-12 208-222 PR00049D 0.00 7.655e-11 202-216 PR00049D 0.00 7.958e-11 204-218 PR00049D 0.00 8.336e-11 203-217 PR00049D 0.00 1.214e-10 209-223 PR00049D 0.00 1.214e-10 210-224 PR00049D 0.00 3.746e-09 211-225
497	PD02059	CORE POLYPROTEIN PROTEIN GAG CONTAINS: P.	PD02059B 24.48 5.056e-09 194-228
497	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 7.706e-11 193-225 DM00215 19.43 5.018e-10 195-227 DM00215 19.43 5.982e-10 192-224 DM00215 19.43 7.750e-10 188-220 DM00215 19.43 7.911e-10 198-230 DM00215 19.43 9.839e-10 189-221 DM00215 19.43 5.271e-09 191-223
497	BL00904	Protein prenyltransferases alpha subunit repeat proteins proteins.	BL00904A 8.30 4.766e-09 205-254 BL00904A 8.30 7.766e-09 204-253
501	BL01113	C1q domain proteins.	BL01113A 17.99 3.106e-10 22-48
501	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513D 11.06 8.085e-09 50-67
502	PR00828	FORMIN SIGNATURE	PR00828H 8.87 4.081e-09 390-411
504	PR00169	POTASSIUM CHANNEL SIGNATURE	PR00169H 8.09 5.696e-30 225-251 PR00169E 9.10 8.773e-28 127-153 PR00169G 9.39 6.684e-27 196-218 PR00169C 16.31 8.714e-25 59-82 PR00169F 7.19 6.192e-24 156-179 PR00169D 12.86 2.385e-20 85-105
507	PR00451	CHITIN-BINDING DOMAIN SIGNATURE	PR00451A 6.49 1.871e-09 88-96
507	PR00873	ECHINOIDEA (SEA URCHIN) METALLOTHIONEIN SIGNATURE	PR00873D 8.43 9.707e-09 78-96
511	PF01327	Polypeptide deformylase.	PF01327D 18.82 2.440e-20 197-228 PF01327A 18.58 2.187e-09 92-126
512	PD02796	PROTEIN STEROL CARRIER LIPID-TRAN.	PD02796B 20.92 6.507e-23 157-203
513	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232A 27.72 7.218e-12 38-70
516	BL00261	Glycoprotein hormones beta chain proteins.	BL00261B 25.64 1.000e-40 72-115 BL00261A 23.97 3.500e-34 22-55
517	PR00796	VIRAL SPIKE GLYCOPROTEIN PRECURSOR SIGNATURE	PR00796I 8.96 7.638e-11 32-57
520	PR00209	ALPHA/BETA GLIADIN FAMILY SIGNATURE	PR00209B 4.88 8.594e-09 129-147
523	PR00833	POLLEN ALLERGEN POA	PR00833H 2.30 6.625e-10 61-75

Table 3
198

SEQ ID NO:	Database entry ID	Description	*Results
		PI SIGNATURE	
523	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308C 3.83 5.846e-10 66-75 PR00308C 3.83 9.308e-10 58-67 PR00308A 5.90 3.859e-09 63-77
523	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 8.685e-11 73-87 PR00456E 3.06 7.375e-10 64-78 PR00456E 3.06 7.844e-10 61-75 PR00456E 3.06 9.625e-10 57-71 PR00456E 3.06 9.625e-10 58-72 PR00456E 3.06 9.625e-10 59-73 PR00456E 3.06 9.906e-10 60-74 PR00456E 3.06 1.228e-09 62-76 PR00456E 3.06 2.367e-09 56-70 PR00456E 3.06 2.595e-09 67-81 PR00456E 3.06 3.962e-09 68-82 PR00456E 3.06 5.443e-09 50-64
523	PF00761	Polyomavirus coat protein.	PF00761B 18.21 6.924e-09 51-89
523	DM01283	A-BINDING PROTEIN CHLOROPHYLL.	DM01283A 14.91 5.300e-10 55-90 DM01283A 14.91 5.781e-09 53-88 DM01283A 14.91 8.313e-09 50-85
542	PR00779	INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN RECEPTOR SIGNATURE	PR00779H 8.81 6.909e-09 18-39
544	DM00031	IMMUNOGLOBULIN V REGION.	DM00031B 15.41 4.508e-15 84-117
549	PD01736	PROTEIN TRANSMEMBRANE INTERGENIC REGION RECQ-PLD.	PD01736B 8.42 9.250e-09 118-129
551	PF00512	Signal carboxyl-terminal domain proteins.	PF00512 13.94 3.571e-14 150-168
552	PF01032	FecCD transport family.	PF01032B 9.12 7.300e-15 132-146
553	BL00713	Sodium:dicarboxylate symporter family proteins.	BL00713D 20.98 6.063e-09 24-61
554	DM00784	APILLOMAVIRUS E4 PROTEIN.	DM00784B 17.87 7.492e-09 67-91
554	PF00624	Flocculin repeat proteins.	PF00624J 6.21 2.669e-10 49-103 PF00624G 10.91 7.225e-10 86-140 PF00624G 10.91 2.016e-09 78-132 PF00624G 10.91 3.831e-09 30-84 PF00624F 11.04 3.976e-09 67-102 PF00624G 10.91 4.339e-09 60-114 PF00624F 11.04 5.355e-09 73-108 PF00624F 11.04 5.935e-09 19-54 PF00624G 10.91 6.589e-09 84-138 PF00624G 10.91 6.734e-09 62-116 PF00624G 10.91 7.677e-09 38-92 PF00624G 10.91 8.403e-09 21-75 PF00624J 6.21 9.023e-09 61-115 PF00624J 6.21 9.023e-09 65-119 PF00624G 10.91 9.347e-09 24-78 PF00624G 10.91 9.710e-09 92-146
559	BL00590	LIF / OSM family proteins.	BL00590B 17.36 3.045e-19 183-200
562	BL00713	Sodium:dicarboxylate symporter family proteins.	BL00713C 19.76 1.964e-09 100-138

Table 3
199

SEQ ID NO:	Database entry ID	Description	*Results
563	BL00216	Sugar transport proteins.	BL00216B 27.64 8.000e-25 108-157
563	PR00171	SUGAR TRANSPORTER SIGNATURE	PR00171C 10.97 8.714e-13 268-278 PR00171D 12.76 1.610e-11 357-378 PR00171B 14.73 2.019e-09 109-128
563	PR00172	GLUCOSE TRANSPORTER SIGNATURE	PR00172A 9.82 9.372e-20 258-279 PR00172F 8.47 6.400e-15 420-440 PR00172B 8.42 7.639e-14 295-316 PR00172E 8.29 5.755e-13 390-408 PR00172D 9.13 2.227e-12 357-380 PR00172C 9.51 2.209e-09 326-346
563	PR00593	METABOTROPIC GLUTAMATE RECEPTOR SIGNATURE	PR00593E 11.51 5.227e-09 112-126
565	BL00979	G-protein coupled receptors family 3 proteins.	BL00979M 14.39 5.114e-12 126-176
565	PR00248	METABOTROPIC GLUTAMATE GPCR SIGNATURE	PR00248F 14.25 8.222e-09 152-174
566	BL00402	Binding-protein-dependent transport systems inner membrane co.	BL00402A 5.93 7.000e-09 55-68
568	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 8.342e-09 24-48
568	PR00175	SODIUM/ALANINE SYMPORTER SIGNATURE	PR00175C 11.57 9.753e-09 2-21
587	PR00170	SODIUM CHANNEL SIGNATURE	PR00170G 7.74 3.374e-09 37-65
594	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 5.974e-12 83-122
594	PR00534	MELANOCORTIN RECEPTOR FAMILY SIGNATURE	PR00534A 11.49 6.123e-10 44-56
594	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 4.484e-17 231-246 PR00245A 18.03 9.265e-16 52-73 PR00245B 10.38 9.514e-12 170-184 PR00245D 10.47 2.465e-10 267-278 PR00245E 12.40 8.302e-10 284-298
594	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237E 13.03 2.800e-10 192-215 PR00237A 11.48 5.935e-09 19-43
605	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 1.419e-18 57-78
605	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 5.875e-11 24-48
606	PR00927	ADENINE NUCLEOTIDE TRANSLOCATOR 1 SIGNATURE	PR00927A 7.98 9.667e-09 14-26
609	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 2.250e-09 58-79 PR00237G 19.63 9.372e-09 143-169
610	PR00698	C.ELEGANS SRG FAMILY INTEGRAL MEMBRANE PROTEIN SIGNATURE	PR00698E 14.43 8.714e-09 97-122
615	PF00075	RNase H.	PF00075A 14.44 4.429e-09 231-247

Table 3
200

SEQ ID NO:	Database entry ID	Description	*Results
618	PF01325	Iron dependant repressor.	PF01325B 20.91 5.680e-09 34-55
619	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 9.727e-36 58-96
620	PR00907	THROMBOMODULIN SIGNATURE	PR00907E 11.70 2.969e-10 49-71
632	PD01115	PRECURSOR AMPHIBIAN SKIN SIGNAL.	PD01115A 12.27 9.750e-12 1-23
636	BL00970	Nuclear transition protein 2 proteins.	BL00970B 10.09 8.966e-10 83-108
638	PF01007	Inward rectifier potassium channel.	PF01007B 17.48 1.000e-08 95-138
654	BL00948	Ribosomal protein S7e proteins.	BL00948A 14.13 5.034e-20 68-90
658	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 4.150e-10 70-83 PR00019B 11.36 9.100e-10 94-107 PR00019A 11.19 8.000e-09 73-86
658	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR00500B 7.74 9.337e-09 178-198
660	BL00476	Fatty acid desaturases family 1 proteins.	BL00476B 18.34 4.938e-09 252-295
660	PR00669	INHIBIN ALPHA CHAIN SIGNATURE	PR00669B 8.27 6.488e-09 179-195
665	BL01253	Type I fibronectin domain proteins.	BL01253C 15.89 6.654e-18 78-116
665	PR00018	KRINGLE DOMAIN SIGNATURE	PR00018C 14.30 3.625e-21 82-102 PR00018A 14.52 3.423e-09 36-51
670	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 6.034e-09 7-21
672	PR00591	SOMATOSTATIN RECEPTOR TYPE 5 SIGNATURE	PR00591B 7.56 4.750e-09 117-131

* Results include in order: Accession No., subtype, e-value, and amino acid position of the signature in the corresponding polypeptide

Table 4A
201

SEQ ID NO:	Pfam Model	Description	E-value	Score
340	trypsin	Trypsin	1.9e-06	23.0
345	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	0.002	-5.3
350	ig	Immunoglobulin domain	1.7e-08	32.5
354	KRAB	KRAB box	6.4e-22	86.3
356	p450	Cytochrome P450	8.3e-13	48.0
362	ABC_tran	ABC transporter	0.0016	-23.4
383	neur_chan	Neurotransmitter-gated ion-channel	4.8e-15	54.0
386	BCCT	BCCT family transporter	8.5e-22	85.8
388	Fumarate_red_D		3.4e-64	226.7
391	HAMP		1.1e-11	52.2
404	7tm_2	7 transmembrane receptor (Secretin family)	0.0039	-87.5
410	C1q	C1q domain	2.2e-45	164.2
416	MCT	Monocarboxylate transporter	4.4e-59	209.7
426	7tm_1	7 transmembrane receptor (rhodopsin family)	5.4e-22	72.0
435	EGF	EGF-like domain	0.00021	28.1
437	DUF6	Integral membrane protein DUF6	0.043	13.8
438	zf-DHHC	DHHC zinc finger domain	1.2e-32	121.9
443	CUB	CUB domain	6.9e-32	119.4
447	Ribosomal_L31e	Ribosomal protein L31e	0.00061	16.6
448	7tm_3	7 transmembrane receptor (metabotropic glutamate family)	0.0073	-95.1
449	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	7.6e-31	115.9
453	7tm_2	7 transmembrane receptor (Secretin family)	3.7e-05	-46.4
455	tsp_1	Thrombospondin type 1 domain	0.028	12.1
473	7tm_1	7 transmembrane receptor (rhodopsin family)	9.3e-40	128.4
474	PDZ	PDZ domain (Also known as DHR or GLGF).	2.1e-42	154.3
477	trypsin	Trypsin	9.8e-99	313.5
484	Peptidase_M1	Peptidase family M1	3.7e-11	32.8
487	ig	Immunoglobulin domain	1.2e-06	26.5
496	SCP	SCP-like extracellular protein	2.9e-21	80.4
501	C1q	C1q domain	5.4e-08	35.2

Table 4A
202

SEQ ID NO:	Pfam Model	Description	E-value	Score
504	ion_trans	Ion transport protein	3.9e-31	116.9
511	Pep_deformylase	Polypeptide deformylase	2.1e-20	81.2
512	SCP2	SCP-2 sterol transfer family	5.2e-23	89.9
513	cadherin	Cadherin domain	2.9e-08	40.9
516	Cys_knot	Cystine-knot domain	3.3e-52	186.9
544	ig	Immunoglobulin domain	2.6e-09	35.1
551	HAMP		1.1e-08	42.3
552	FecCD_family	FecCD transport family	7.4e-44	159.1
553	BPD_transp	Binding-protein-dependent transport systems inner membrane component	6e-05	29.9
557	ig	Immunoglobulin domain	8.8e-13	46.2
559	LIF_OSM	LIF / OSM family	8e-145	494.5
562	SDF	Sodium:dicarboxylate symporter family	3.4e-58	206.8
563	sugar_tr	Sugar (and other) transporter	2e-99	343.7
565	7tm_3	7 transmembrane receptor (metabotropic glutamate family)	2.1e-06	-21.8
576	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	4.1e-08	40.4
579	zf-DHHC	DHHC zinc finger domain	0.0085	-6.4
582	Rhomboid	Rhomboid family	0.072	-20.3
592	ig	Immunoglobulin domain	1.5e-05	23.0
594	7tm_1	7 transmembrane receptor (rhodopsin family)	7.1e-30	97.1
605	7tm_1	7 transmembrane receptor (rhodopsin family)	3.8e-06	21.7
609	7tm_1	7 transmembrane receptor (rhodopsin family)	0.064	8.3
611	DUF6	Integral membrane protein DUF6	1.4e-05	32.0
615	rvt	Reverse transcriptase (RNA-dependent DNA polymerase)	3e-15	61.0
619	KRAB	KRAB box	2e-42	154.4
632	Gastrin	Gastrin/cholecystokinin family	7.5e-22	83.9
634	Cornifin		0.0031	5.4
638	ion_trans	Ion transport protein	0.0034	24.0
642	Galactosyl_T	Galactosyltransferase	2.9e-28	107.3
654	Ribosomal_S7e	Ribosomal protein S7e	6.9e-17	69.5
658	LRR	Leucine Rich Repeat	1.8e-15	64.8
665	kringle	Kringle domain	1.2e-17	72.1

Table 4A
203

SEQ ID NO:	Pfam Model	Description	E-value	Score
666	p450	Cytochrome P450	0.034	10.6

Table 4B
204

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
345	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	0.002	-5.3	1	4-195
350	ig	Immunoglobulin domain	4.3e-05	30.3	1	35-112
351	LRR	Leucine Rich Repeat	1.5	15.3	1	19-41
354	KRAB	KRAB box	3.9e-23	90.3	1	127-167
357	MOSC_N	MOSC N-terminal beta barrel domain	0.00046	11.9	1	54-165
358	LRRNT	Leucine rich repeat N-terminal domain	0.28	17.4	1	35-62
360	Adeno_E3_CR2	Adenovirus E3 region protein CR2	3.9	-1.3	1	83-130
362	ABC_tran	ABC transporter	0.0068	-28.4	1	155-257
381	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	7.5	-67.1	1	1-134
383	Neur_chan_memb	Neurotransmitter-gated ion-channel tra	0.28	-95.0	1	5-100
385	Vps26	Vacuolar protein sorting-associated protein	1.4e-92	321.0	1	12-247
386	Transposase_27	IS1 transposase	1.2e-52	188.3	1	212-323
386	BCCT	BCCT family transporter	8.5e-22	85.8	1	17-328
388	Fumarate_red_D	Fumarate reductase subunit D	1e-63	225.1	1	2-119
390	wzz	Chain length determinant protein	0.003	2.8	1	1-101
391	HAMP	HAMP domain	1.2e-11	52.2	1	74-143
391	LEA	Late embryogenesis abundant protein	3.5	-2.1	1	149-213
404	7tm_2	7 transmembrane receptor (Secretin family)	0.004	-87.9	1	13-153
409	DAG_PE-bind	Phorbol esters/diacylglycerol binding dom	0.73	-6.1	1	59-94
410	C1q	C1q domain	5.9e-46	166.1	1	73-202
416	FecCD	FecCD transport family	0.65	-198.0	1	3-224
416	UPF0118	Domain of unknown function DUF20	2.8	-117.6	1	4-353
416	sugar_tr	Sugar (and other) transporter	6.7	-193.3	1	3-355
416	TerC	Integral membrane protein TerC family	8.9	-103.2	1	26-215
416	secY	eubacterial secY protein	9.2	-248.7	1	5-302
422	MCPsignal	Methyl-accepting chemotaxis protein (MCP) s	0.86	-122.1	1	265-467
422	LEA	Late embryogenesis abundant protein	6.5	-5.5	1	401-463
426	7tm_1	7 transmembrane receptor (rhodopsin family)	0.00024	-17.2	1	18-226
435	EGF	EGF-like domain	0.00021	28.1	5	27-51:64-94:107-137:150-180:193-217
435	EB	EB module	2.8	-6.5	1	113-180

Table 4B
205

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
435	laminin_EGF	Laminin EGF-like (Domains III and V)	3.6	-9.8	4	28-64:68-107:111-150:154-197
437	DUF6	Integral membrane protein DUF6	0.08	10.6	1	160-288
437	DUF250	Domain of unknown function, DUF250	9.9	-107.0	1	142-274
438	zf-DHHC	DHHC zinc finger domain	1.2e-32	121.9	1	98-162
438	CDP-OH_P_transf	CDP-alcohol phosphatidyltransferase	9.9	-35.1	1	14-188
440	DUF6	Integral membrane protein DUF6	0.59	-4.1	1	133-264
443	CUB	CUB domain	4e-31	116.8	1	135-240
443	sushi	Sushi domain (SCR repeat)	4.5e-06	33.6	1	74-131
447	Ribosomal_L31e	Ribosomal protein L31e	0.00061	16.6	1	77-143
448	7tm_3	7 transmembrane receptor	0.0073	-95.1	1	1-108
449	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	7.6e-31	115.9	1	4-181
453	7tm_2	7 transmembrane receptor (Secretin family)	3.7e-05	-46.4	1	1-175
455	tsp_1	Thrombospondin type 1 domain	0.028	12.1	1	31-83
460	Brevenin	Brevenin/esculentin/gaeguri n/rugosin family	8.5	-3.3	1	17-61
464	Pep_M12B_prop ep	Reprolysin family propeptide	0.12	-23.7	1	179-288
472	TMS_TDE	TMS membrane protein/tumour differentially e	2e-06	-157.8	1	1-193
472	SPW	SPW repeat	8.5	-8.3	1	69-121
473	7tm_1	7 transmembrane receptor (rhodopsin family)	1.2e-32	121.9	1	21-279
474	PDZ	PDZ domain (Also known as DHR or GLGF)	1.1e-41	152.0	3	119-211:233-313:314-401
474	Autoind_bind	Autoinducer binding domain	9.3	-50.4	1	26-153
477	trypsin	Trypsin	8.2e-91	315.1	1	36-258
487	ig	Immunoglobulin domain	0.0017	25.1	1	32-109
496	SCP	SCP-like extracellular protein	1.5e-16	68.4	1	18-215
502	MBOAT	MBOAT family	7.5e-73	255.4	1	66-379
504	ion_trans	Ion transport protein	1.4e-32	121.7	1	56-247
504	oxidored_q3	NADH-ubiquinone/plastoquinone oxidoreduct	5.6	-81.7	1	92-242
510	HEAT	HEAT repeat	0.39	17.2	1	106-143
511	Pep_deformylase	Polypeptide deformylase	4.3e-19	76.8	1	63-238
512	SCP2	SCP-2 sterol transfer family	5.2e-23	89.9	1	100-208
512	Uteroglobin	Uteroglobin family	5.6	-26.8	1	1-70
513	cadherin	Cadherin domain	1e-08	42.4	1	48-139

Table 4B
206

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
516	Cys_knot	Cystine-knot domain	8.3e-53	188.9	1	15-125
542	LAG1	Longevity-assurance protein (LAG1)	3.4	-109.3	1	60-243
543	NIF	NLI interacting factor	5.3e-15	63.3	1	120-300
544	ig	Immunoglobulin domain	1.8e-07	38.2	1	34-117
550	YjgP_YjgQ	Predicted permease YjgP/YjgQ family	6.5e-23	89.6	1	1-279
550	Hexose_dehydrat	NDP-hexose 2,3-dehydratase	9.5	-172.0	1	25-147
551	HAMP	HAMP domain	1.1e-08	42.3	1	70-138
551	signal	His Kinase A (phosphoacceptor) domain	1.5	-1.9	1	142-174
552	FecCD	FecCD transport family	7.4e-44	159.1	1	1-203
552	ABC-3	ABC 3 transport family	3.1	-186.2	1	1-203
553	BPD_transp	Binding-protein-dependent transport system	6e-05	29.9	1	108-184
553	Competence	Competence protein	0.71	-84.6	1	2-216
557	ig	Immunoglobulin domain	4.3e-07	37.0	1	45-164
559	LIF_OSM	LIF / OSM family	8e-145	494.5	1	2-209
562	SDF	Sodium:dicarboxylate symporter family	8.3e-07	-83.3	1	1-173
563	sugar_tr	Sugar (and other) transporter	2.1e-99	343.6	1	17-455
563	OATP_C	Organic Anion Transporter Polypeptide	5	-230.7	1	14-348
563	Nuc_H_symport	Nucleoside H+ symporter	5.5	-269.8	1	38-445
563	COX1	Cytochrome C and Quinol oxidase polyp	5.6	-307.7	1	9-422
563	DUF21	Domain of unknown function DUF21	7.2	-75.4	1	22-207
563	PUCC	PUCC protein	7.4	-280.0	1	37-444
563	xan_ur_permease	Permease family	9.7	-202.9	1	5-349
563	DUF318	Predicted permease	10	-169.2	1	82-367
565	7tm_3	7 transmembrane receptor	2.1e-06	-22.0	1	1-184
570	DUF323	Domain of unknown function (DUF323)	0.0018	-58.7	1	31-150
576	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	4.1e-08	40.4	1	7-183
579	zf-DHHC	DHHC zinc finger domain	0.0085	-6.4	1	5-40
582	Rhomboid	Rhomboid family	0.092	-22.0	1	15-98
587	ion_trans	Ion transport protein	0.18	10.6	1	2-133
592	ig	Immunoglobulin domain	0.0031	24.2	1	40-111
594	7tm_1	7 transmembrane receptor (rhodopsin family)	1.9e-27	104.6	1	34-283
603	Folate_rec	Folate receptor family	0.87	-107.5	1	6-212
611	DUF6	Integral membrane protein DUF6	0.00017	28.3	2	8-129:147-277
611	PhaG_MnhG_YufB	Na+/H+ antiporter subunit	2	-50.3	1	16-118
611	DUF7	Integral membrane protein DUF7	3.9	-34.6	1	177-268
611	Competence	Competence protein	7.5	-104.9	1	43-280

Table 4B
207

SEQ ID NO:	Pfam Model	Description	E-value	Score	No: of Pfam Domains	Position of the Domain
615	rvt	Reverse transcriptase	1.5e-08	41.9	1	214-381
619	KRAB	KRAB box	6.4e-27	102.9	1	56-96
621	MAPEG	MAPEG family	2.1	-21.7	1	10-95
632	Gastrin	Gastrin/cholecystokinin family	4e-05	30.5	1	2-74
634	Cornifin	Cornifin (SPRR) family	0.0031	5.4	1	8-221
638	ion_trans	Ion transport protein	0.01	22.4	1	101-263
642	Galactosyl_T	Galactosyltransferase	1.2e-25	98.6	1	130-334
653	DUF312	Short repeats of unknown function (DUF312)	9.2	-2.8	1	269-312
654	Ribosomal_S7e	Ribosomal protein S7e	1e-16	69.0	1	66-158
658	LRR	Leucine Rich Repeat	1.2e-15	65.5	5	48-71:72-95:96-119:120-143:144-167
658	LRRNT	Leucine rich repeat N-terminal domain	3e-08	40.9	1	17-46
658	LRRCT	Leucine rich repeat C-terminal domain	7.8e-07	36.1	1	177-230
665	kringle	Kringle domain	1.2e-17	72.1	1	36-119
665	CUB	CUB domain	2.5e-12	54.4	1	219-323
665	WSC	WSC domain	2.6e-08	41.0	1	124-205
668	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	8.6	-68.2	1	25-200

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
350	1a14	L	20	126	3.4e-25			55.31	NEURAMINIDASE; CHAIN: N; SINGLE CHAIN ANTIBODY; CHAIN: H, L;	COMPLEX (ANTIBODY/ANTIGEN) COMPLEX (ANTIBODY/ANTIGEN), SINGLE-CHAIN ANTIBODY, 2 GLYCOSYLATED PROTEIN
350	1a2y	A	20	126	5.1e-27			54.70	MONOCLONAL ANTIBODY D1.3; CHAIN: A, B; LYSOZYME; CHAIN: C;	COMPLEX (IMMUNOGLOBULIN/HYDROLASE) COMPLEX (IMMUNOGLOBULIN/HYDROLASE), IMMUNOGLOBULIN V 2 REGION, SIGNAL, HYDROLASE, GLYCOSIDASE, BACTERIOLYTIC 3 ENZYME, EGG WHITE
350	1a7q	L	20	136	1.5e-25			52.86	MONOCLONAL ANTIBODY D1.3; CHAIN: L, H;	IMMUNOGLOBULIN IMMUNOGLOBULIN, VARIANT
350	1ao7	E	22	142	3.4e-46	-0.08	0.06		HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA-A2 HEAVY CHAIN; CLASS I MHC, T-CELL RECEPTOR, VIRAL PEPTIDE, 2 COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR
350	1ap2	A	20	128	5.1e-30			51.96	MONOCLONAL ANTIBODY C219; CHAIN: A, B, C, D;	IMMUNOGLOBULIN VARIABLE DOMAIN; SINGLE CHAIN FV, MONOCLONAL ANTIBODY, C219, P-GLYCOPROTEIN, 2 IMMUNOGLOBULIN
350	1ar1	D	20	136	3.4e-26			52.90	CYTOCHROME C OXIDASE; CHAIN: A, B; ANTIBODY FV FRAGMENT; CHAIN: C, D;	COMPLEX (OXIDOREDUCTASE/ANTIBODY) CYTOCHROME AA3, COMPLEX

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										IV, FERROCYTOCHROME C, COMPLEX (OXIDOREDUCTASE/ANTIBODY), ELECTRON TRANSPORT, 2 TRANSMEMBRANE, CYTOCHROME OXIDASE, ANTIBODY COMPLEX
350	1b0w	A	20	127	5.1e-27			55.43	BENCE-JONES KAPPA I PROTEIN BRE; CHAIN: A, B, C;	IMMUNE SYSTEM BENCE-JONES; IMMUNOGLOBULIN, AMYLOID, IMMUNE SYSTEM
350	1bd2	E	22	160	1.7e-48	-0.10	0.07		HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA A2 HEAVY CHAIN; COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR)
350	1bec		23	143	1.7e-46	0.27	0.30		14.3.D T CELL ANTIGEN RECEPTOR; 1BEC 5 CHAIN: NULL; 1BEC 6	RECEPTOR T CELL RECEPTOR 1BEC 14
350	1bfv	L	20	127	1.7e-25			51.18	FV4155; CHAIN: L, H;	IMMUNOGLOBULIN IMMUNOGLOBULIN, FV FRAGMENT, STEROID HORMONE, 2 FINE SPECIFICITY
350	1bvk	A	20	127	1.2e-29			57.64	HULYS11; CHAIN: A, B, D, E; LYSOZYME; CHAIN: C, F;	COMPLEX (HUMANIZED ANTIBODY/HYDROLASE) MURAMIDASE; HUMANIZED ANTIBODY, ANTIBODY COMPLEX, FV, ANTI-LYSOZYME, 2 COMPLEX (HUMANIZED ANTIBODY/HYDROLASE)

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
350	1bwm	A	23	138	3.4e-45	0.06	0.12		ALPHA-BETA T CELL RECEPTOR (TCR) (D10); CHAIN: A;	IMMUNE SYSTEM IMMUNOGLOBULIN, IMMUNORECEPTOR, IMMUNE SYSTEM
350	1bww	A	18	126	1e-28			52.26	IG KAPPA CHAIN V-I REGION REI; CHAIN: A, B;	IMMUNE SYSTEM REIV, STABILIZED IMMUNOGLOBULIN FRAGMENT, BENCE-JONES 2 PROTEIN, IMMUNE SYSTEM
350	1d9k	B	23	138	3.4e-45	0.12	0.34		T-CELL RECEPTOR D10 (ALPHA CHAIN); CHAIN: A, E; T-CELL RECEPTOR D10 (BETA CHAIN); CHAIN: B, F; MHC I-AK A CHAIN (ALPHA CHAIN); CHAIN: C, G; MHC I-AK B CHAIN (BETA CHAIN); CHAIN: D, H; CONALBUMIN PEPTIDE; CHAIN: P, Q;	IMMUNE SYSTEM MHC I-AK; MHC I-AK; T-CELL RECEPTOR, MHC CLASS II, D10, I-AK
350	1dlf	L	20	127	8.5e-26			54.31	ANTI-DANSYL IMMUNOGLOBULIN IGG2A(S); CHAIN: L, H;	IMMUNOGLOBULIN ANTI-DANSYL FV FRAGMENT FV FRAGMENT, IMMUNOGLOBULIN
350	1dsf	L	20	129	5.1e-23			53.77	ANTICANCER ANTIBODY B1; CHAIN: L, H;	IMMUNOGLOBULIN B1DSFV; MONOCLONAL ANTIBODY, ANTITUMOR, IMMUNOGLOBULIN
350	1f11	A	20	159	6.8e-34	-0.00	0.07		F124 IMMUNOGLOBULIN (KAPPA LIGHT CHAIN); CHAIN: A, C; F124 IMMUNOGLOBULIN (IGG1 HEAVY CHAIN); CHAIN: B, D;	IMMUNE SYSTEM IMMUNOGLOBULIN, ANTIBODY, FAB, HEPATITIS B, PRES2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
350	1fgv	L	20	136	1.7e-31			55.11	IMMUNOGLOBULIN FV FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 1FGV 3 ANTIBODY 'H52' (HUH52-AA FV) 1FGV 4	
350	1fvc	A	20	128	6.8e-31			54.26	IMMUNOGLOBULIN FV FRAGMENT OF HUMANIZED ANTIBODY 4D5, VERSION 8 1FVC 3	
350	1fyt	E	22	160	6.8e-44	0.05	0.10		HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR CHAIN: A; HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR-1 CHAIN: B; HEMAGGLUTININ HA1 PEPTIDE CHAIN; CHAIN: C; T-CELL RECEPTOR ALPHA CHAIN; CHAIN: D; T-CELL RECEPTOR BETA CHAIN; CHAIN: E;	IMMUNE SYSTEM HLA-DR1, DRA; HLA-DR1, DRB1 0101; TCR HA1.7 ALPHA CHAIN; TCR HA1.7 BETA CHAIN; PROTEIN-PROTEIN COMPLEX, IMMUNOGLOBULIN FOLD
350	1igm	L	20	134	5.1e-30			57.34	IMMUNOGLOBULIN IMMUNOGLOBULIN M (IG-M) FV FRAGMENT 1IGM 3	
350	1ivl	A	20	126	1e-24			60.97	IMMUNOGLOBULIN IMMUNOGLOBULIN VL DOMAIN (VARIABLE DOMAIN OF KAPPA LIGHT 1IVL 3 CHAIN) OF DESIGNED ANTIBODY M29B 1IVL 4	
350	1jhl	L	20	127	3.4e-28			57.95	COMPLEX(ANTIBODY-	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									ANTIGEN) FV FRAGMENT (IGG1, KAPPA) (LIGHT AND HEAVY VARIABLE DOMAINS IJHL 3 NON-COVALENTLY ASSOCIATED) OF MONOCLONAL ANTI-HEN EGG IJHL 4 LYSOZYME ANTIBODY D11.15 COMPLEX WITH PHEASANT EGG IJHL 5 LYSOZYME IJHL 6	
350	1kb5	B	21	136	1.7e-33			50.75	KB5-C20 T-CELL ANTIGEN RECEPTOR; CHAIN: A, B; ANTIBODY DESIRE-1; CHAIN: L, H;	COMPLEX (IMMUNOGLOBULIN/RECEPTOR) TCR VAPLHA VBETA DOMAIN; T-CELL RECEPTOR, STRAND SWITCH, FAB, ANTICLONOTYPIC, 2 (IMMUNOGLOBULIN/RECEPTOR)
350	1maj		20	127	6.8e-24			50.59	IMMUNOGLOBULIN MURINE ANTIBODY 26-10 VL DOMAIN (NMR, 15 ENERGY MINIMIZED IMAJ 3 STRUCTURES) IMAJ 4	
350	1nfd	B	20	143	1e-45	0.08	0.27		N15 ALPHA-BETA T-CELL RECEPTOR; CHAIN: A, B, C, D; H57 FAB; CHAIN: E, F, G, H	COMPLEX (IMMUNORECEPTOR/IMMUNOG LOBULIN) COMPLEX (IMMUNORECEPTOR/IMMUNOG LOBULIN)
350	1nmb	L	20	128	8.5e-27			58.89	N9 NEURAMINIDASE; INMB 4 CHAIN: N; INMB 5 FAB NC10; INMB 9 CHAIN: L, H; INMB 10	COMPLEX (HYDROLASE/IMMUNOGLOBUL IN)
350	1rvf	L	20	130	5.1e-26			54.02	HUMAN RHINOVIRUS 14	COMPLEX (COAT

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									COAT PROTEIN; CHAIN: 1, 2, 3, 4; FAB 17-IA; CHAIN: L, H	PROTEIN/IMMUNOGLOBULIN) POLYPROTEIN, COAT PROTEIN, CORE PROTEIN, RNA-DIRECTED RNA 2 POLYMERASE, HYDROLASE, THIOL PROTEASE, MYRISTYLATION, 3 COMPLEX (COAT
										PROTEIN/IMMUNOGLOBULIN)
350	1sbs	L	20	159	1.2e-33	0.10	0.33		MONOCLONAL ANTIBODY 3A2; CHAIN: H, L;	MONOCLONAL ANTIBODY MONOCLONAL ANTIBODY, FAB-FRAGMENT, REPRODUCTION
350	1tcr	B	20	143	1e-45	0.06	0.17		ALPHA, BETA T-CELL RECEPTOR CHAIN: A, B;	RECEPTOR TCR; T-CELL, RECEPTOR, TRANSMEMBRANE, GLYCOPROTEIN, SIGNAL
350	1wtl	A	20	127	6.8e-28		54.08		IMMUNOGLOBULIN WAT, A VARIABLE DOMAIN FROM IMMUNOGLOBULIN LIGHT-CHAIN 1WTL 3 (BENCE-	
350	2rhe		21	130	1.7e-24		52.52		IMMUNOGLOBULIN BENCE-*JONES PROTEIN (LAMBDA, VARIABLE DOMAIN) 2RHE 4	
351	1fo1	A	1	53	0.00012	-0.34	0.12		NUCLEAR RNA EXPORT FACTOR 1; CHAIN: A, B;	RNA BINDING PROTEIN TAP (NFX1); RIBONUCLEOPROTEIN (RNP,RBD OR RRM) AND LEUCINE-RICH-REPEAT 2 (LRR)
356	1dt6	A	60	248	8.5e-52	-0.41	0.05		CYTOCHROME P450 2C5; CHAIN: A;	OXIDOREDUCTASE PROGESTERONE 21-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										HYDROXYLASE, CYPIIC5 P450 1, MEMBRANE PROTEIN, PROGESTERONE 21- HYDROXYLASE, BENZO(A) 2 PYRENE HYDROXYLASE, ESTRADIOL 2-HYDROXYLASE, P450, CYP2C5
362	1b0u	A	129	254	5.1e-24	0.37	0.36		HISTIDINE PERMEASE; CHAIN: A;	TRANSPORT PROTEIN ABC TRANSPORTER, HISP; ABC TRANSPORTER, HISTIDINE PERMEASE, TRANSPORT PROTEIN
362	1f2u	A	141	175	0.0025	-0.78	0.09		RAD50 ABC-ATPASE; CHAIN: A, C; RAD50 ABC-ATPASE; CHAIN: B, D;	REPLICATION DNA DOUBLE-STRAND BREAK REPAIR, ABC-ATPASE
362	1f2u	A	160	213	0.0048	-0.91	0.12		RAD50 ABC-ATPASE; CHAIN: A, C; RAD50 ABC-ATPASE; CHAIN: B, D;	REPLICATION DNA DOUBLE-STRAND BREAK REPAIR, ABC-ATPASE
362	1g29	1	142	253	3.4e-21	-0.28	0.34		MALTOSE TRANSPORT PROTEIN MALK; CHAIN: 1, 2;	SUGAR BINDING PROTEIN MALK; ATPASE, ACTIVE TRANSPORT, MALTOSE UPTAKE AND REGULATION
362	1gky		158	184	0.0027	-0.82	0.28		TRANSFERASE GUANYLATE KINASE (E.C.2.7.4.8) COMPLEX WITH 1GKY 3 GUANOSINE MONOPHOSPHATE 1GKY 4	
364	1e3y	A	103	159	0.0025	0.21	0.52		FADD PROTEIN; CHAIN: A;	APOPTOSIS FAS-ASSOCIATING DEATH DOMAIN-CONTAINING

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										PROTEIN ; DEATH DOMAIN, ADAPTER MOLECULE, FAS RECEPTOR DEATH INDUCING 2 SIGNALLING COMPLEX
364	1fad	A	103	150	0.00075	0.14	1.00		FADD PROTEIN; CHAIN: A;	APOPTOSIS APOPTOSIS, FADD, DEATH DOMAIN
364	1lrv		19	202	0.0018	0.32	0.03		LEUCINE-RICH REPEAT VARIANT; CHAIN: NULL;	LEUCINE-RICH REPEATS LRV; LEUCINE-RICH REPEATS, REPETITIVE STRUCTURE, IRON SULFUR 2 PROTEINS, NITROGEN FIXATION
388	1fum	D	2	100	1.7e-44	-0.68	1.00		FUMARATE REDUCTASE FLAVOPROTEIN SUBUNIT; CHAIN: A, M; FUMARATE REDUCTASE IRON-SULFUR PROTEIN; CHAIN: B, N; FUMARATE REDUCTASE 15 KD HYDROPHOBIC PROTEIN; CHAIN: C, O; FUMARATE REDUCTASE 13 KD HYDROPHOBIC PROTEIN; CHAIN: D, P;	OXIDOREDUCTASE COMPLEX II; COMPLEX II; COMPLEX II; COMPLEX II; FUMARATE REDUCTASE, COMPLEX II, SUCCINATE DEHYDROGENASE, 2 RESPIRATION, OXIDOREDUCTASE
388	1fum	D	2	117	1.7e-44			168.49	FUMARATE REDUCTASE FLAVOPROTEIN SUBUNIT; CHAIN: A, M; FUMARATE REDUCTASE IRON-SULFUR PROTEIN; CHAIN: B, N; FUMARATE REDUCTASE 15 KD HYDROPHOBIC PROTEIN; CHAIN: C, O; FUMARATE REDUCTASE 13 KD	OXIDOREDUCTASE COMPLEX II; COMPLEX II; COMPLEX II; COMPLEX II; FUMARATE REDUCTASE, COMPLEX II, SUCCINATE DEHYDROGENASE, 2 RESPIRATION, OXIDOREDUCTASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									HYDROPHOBIC PROTEIN; CHAIN: D, P;	
391	1qu7	A	154	214	2.5e-09	-0.49	0.90		METHYL-ACCEPTING CHEMOTAXIS PROTEIN I; CHAIN: A, B;	SIGNALING PROTEIN SERINE, CHEMOTAXIS, FOUR HELICAL-BUNDLE
391	2asr		38	71	5e-10	-0.81	0.51		CHEMOTAXIS ASPARTATE RECEPTOR (LIGAND BINDING DOMAIN) 2ASR 3	
391	2lig	A	26	71	2.5e-14	-0.79	0.47		ASPARTATE RECEPTOR; 2LIG 4 CHAIN: A, B; 2LIG 5	CHEMOTAXIS
396	1c17	M	130	265	0.001			73.12	ATP SYNTHASE SUBUNIT C; CHAIN: A, B, C, D, E, F, G, H, I, J, K, L; ATP SYNTHASE SUBUNIT A; CHAIN: M;	MEMBRANE PROTEIN MEMBRANE PROTEIN, HELIX, COMPLEX
402	1d0s	A	2	91	1.3e-09	0.26	-0.20		NICOTINATE MONONUCLEOTIDE:5,6-CHAIN: A;	TRANSFERASE DINUCLEOTIDE-BINDING MOTIF, PHOSPHORIBOSYL TRANSFERASE
402	1eut		24	125	1e-09	0.40	-0.20		SIALIDASE; CHAIN: NULL;	HYDROLASE NEURAMINIDASE; HYDROLASE, GLYCOSIDASE
402	2pro	A	10	136	1e-18	0.12	-0.20		ALPHA-LYTIC PROTEASE; CHAIN: A, B, C;	PRO REGION PRO REGION, FOLDASE, PROTEIN FOLDING, SERINE PROTEASE
410	1c28	A	71	204	1.7e-34	0.72	0.89		30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C;	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA, SERUM PROTEIN
410	1c28	A	73	203	6.8e-33	0.52	0.98		30 KD ADIPOCYTE	SERUM PROTEIN ACRP30 C1Q

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C;	TNF TRIMER ALL-BETA, SERUM PROTEIN
410	1c28	A	77	204	1.7e-34			64.45	30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C;	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA, SERUM PROTEIN
410	1c28	B	73	203	1e-30	0.76	0.83		30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C;	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA, SERUM PROTEIN
410	1c28	B	81	196	1e-30			53.80	30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C;	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA, SERUM PROTEIN
410	1c28	C	73	203	8.5e-28	0.56	0.37		30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN CHAIN: A, B, C;	SERUM PROTEIN ACRP30 C1Q TNF TRIMER ALL-BETA, SERUM PROTEIN
414	1bhd	A	42	87	8.5e-18	0.00	0.04		UTROPHIN; CHAIN: A, B;	STRUCTURAL PROTEIN CALPONIN HOMOLOGY, ACTIN BINDING, STRUCTURAL PROTEIN
414	1bkr	A	41	89	1.7e-20	-0.24	0.28		SPECTRIN BETA CHAIN; CHAIN: A;	ACTIN-BINDING CALPONIN HOMOLOGY (CH) DOMAIN; FILAMENTOUS ACTIN-BINDING DOMAIN, CYTOSKELETON
414	1dxx	A	26	76	1e-09	-0.48	0.41		DYSTROPHIN; CHAIN: A, B, C, D;	STRUCTURAL PROTEIN DYSTROPHIN, MUSCULAR DYSTROPHY, CALPONIN HOMOLOGY DOMAIN, 2 ACTIN-BINDING, UTROPHIN
414	1dxx	A	42	89	1.5e-16	-0.35	0.11		DYSTROPHIN; CHAIN: A, B, C, D;	STRUCTURAL PROTEIN DYSTROPHIN, MUSCULAR DYSTROPHY, CALPONIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										HOMOLOGY DOMAIN, 2 ACTIN-BINDING, UTROPHIN
414	1qag	A	42	87	8.5e-18	-0.59	0.09		UTROPHIN ACTIN BINDING REGION; CHAIN: A, B;	STRUCTURAL PROTEIN CALPONIN HOMOLOGY DOMAIN, DOMAIN SWAPPING, ACTIN BINDING, 2 UTROPHIN, DYSTROPHIN, STRUCTURAL PROTEIN
422	4hbl		290	328	0.00051	0.28	0.53		DHP1; CHAIN: NULL;	DESIGNED HELICAL BUNDLE DESIGNED HELICAL BUNDLE
435	1aut	L	97	202	2.5e-13			51.11	ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAJ; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
435	1dan	L	114	245	5e-16			53.38	BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLOROMETHYLKETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
435	1dan	L	151	232	8.5e-12	0.07	0.30		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: T, U; D-PHE-PHE-ARG-CHLOROMETHYLKETONE (DFFRCMK) WITH CHAIN: C;	INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
435	1dan	L	32	154	2.5e-15	0.23	-0.13		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLOROMETHYLKETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
435	1dan	L	82	197	5e-16	0.45	-0.12		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLOROMETHYLKETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
435	1dva	L	151	232	8.5e-12	-0.02	0.63		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
435	1dx5	I	107	225	2.5e-15	0.30	0.04		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										ANTIFIBRINOLYTIC COMPLEX
435	1dx5	I	149	259	6.8e-14	-0.00	-0.18		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2
										ANTIFIBRINOLYTIC COMPLEX
435	1dx5	I	70	193	2e-16	0.42	-0.12		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2
										ANTIFIBRINOLYTIC COMPLEX
435	1ext	A	33	173	5e-15	0.13	-0.15		TUMOR NECROSIS FACTOR RECEPTOR; CHAIN: A, B;	SIGNALLING PROTEIN BINDING PROTEIN, CYTOKINE, SIGNALLING PROTEIN
435	1ext	A	53	203	1.8e-15			65.24	TUMOR NECROSIS FACTOR RECEPTOR; CHAIN: A, B;	SIGNALLING PROTEIN BINDING PROTEIN, CYTOKINE, SIGNALLING PROTEIN
435	1ext	A	54	197	1.8e-15	0.30	-0.06		TUMOR NECROSIS FACTOR RECEPTOR; CHAIN: A, B;	SIGNALLING PROTEIN BINDING PROTEIN, CYTOKINE, SIGNALLING PROTEIN
435	1fak	L	151	232	8.5e-12	-0.08	0.78		BLOOD COAGULATION FACTOR VIIA; CHAIN: L; BLOOD COAGULATION	BLOOD CLOTTING COMPLEX(SERINE PROTEASE/COFACTOR/LIGAND)

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									FACTOR VIIA; CHAIN: H; SOLUBLE TISSUE FACTOR; CHAIN: T; 5L15; CHAIN: I;	, BLOOD COAGULATION, 2 SERINE PROTEASE, COMPLEX, CO-FACTOR, RECEPTOR ENZYME, 3 INHIBITOR, GLA, EGF, COMPLEX (SERINE 4 PROTEASE/COFACTOR/LIGAND) , BLOOD CLOTTING
435	1igr	A	8	214	1e-21	0.15	-0.12		INSULIN-LIKE GROWTH FACTOR RECEPTOR 1; CHAIN: A;	HORMONE RECEPTOR HORMONE RECEPTOR, INSULIN RECEPTOR FAMILY
435	1klo		111	237	3.4e-17	0.52	0.88		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
435	1klo		154	268	8.5e-16	0.34	0.10		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
435	1klo		31	198	2.5e-29	0.39	0.01		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
435	1klo		33	199	2.5e-29			91.67	LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
435	1klo		68	197	1.2e-18	0.51	0.90		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
435	1klo		68	218	2.5e-26	0.24	-0.14		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
435	1ncf	A	39	180	5e-17	0.23	-0.07		TUMOR NECROSIS FACTOR RECEPTOR; 1NCF 4 CHAIN: A, B; 1NCF 5	SIGNALLING PROTEIN TYPE I RECEPTOR, STNFR1; 1NCF 8 BINDING PROTEIN, CYTOKINE 1NCF 19
435	1ncf	A	51	180	5e-17			54.09	TUMOR NECROSIS FACTOR RECEPTOR; 1NCF 4 CHAIN: A, B; 1NCF 5	SIGNALLING PROTEIN TYPE I RECEPTOR, STNFR1; 1NCF 8 BINDING PROTEIN, CYTOKINE 1NCF 19
435	1ncf	A	96	218	2.5e-16	0.33	-0.14		TUMOR NECROSIS FACTOR	SIGNALLING PROTEIN TYPE I

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									RECEPTOR; 1NCF 4 CHAIN: A, B; 1NCF 5	RECEPTOR, STNFR1; 1NCF 8 BINDING PROTEIN, CYTOKINE 1NCF 19
435	1pfx	L	64	214	1.8e-26	0.20	-0.18		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
435	1pfx	L	72	208	5e-28			62.61	FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
435	1pp2	R	64	184	1e-17	0.13	-0.18		HYDROLASE CALCIUM-FREE PHOSPHOLIPASE A=2= (E.C.3.1.1.4) 1PP2 4	
435	1qfk	L	107	206	7.5e-17	0.34	-0.02		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
435	1qfk	L	151	232	8.5e-12	0.04	0.83		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE
435	1qub	A	11	297	2.5e-33			60.56	HUMAN BETA2-GLYCOPROTEIN I; CHAIN: A;	MEMBRANE ADHESION SHORT CONSENSUS REPEAT, SUSHI, COMPLEMENT CONTROL PROTEIN, 2 N-GLYCOSYLATION, MULTI-DOMAIN, MEMBRANE ADHESION
435	1skz		106	216	2.5e-20			66.23	ANTISTASIN; CHAIN: NULL;	SERINE PROTEASE INHIBITOR FACTOR XA INHIBITOR; ANTISTASIN, CRYSTAL STRUCTURE, FACTOR XA INHIBITOR, 2 SERINE PROTEASE INHIBITOR, THROMBOSIS
435	1skz		64	216	2.5e-20	0.18	0.21		ANTISTASIN; CHAIN: NULL;	SERINE PROTEASE INHIBITOR FACTOR XA INHIBITOR; ANTISTASIN, CRYSTAL STRUCTURE, FACTOR XA INHIBITOR, 2 SERINE PROTEASE INHIBITOR, THROMBOSIS
435	1tpg		37	143	2.2e-18	0.38	0.11		T-PLASMINOGEN ACTIVATOR F1-G; ITPG 7 CHAIN: NULL; ITPG 8	PLASMINOGEN ACTIVATION
435	1tpg		81	184	5e-18	0.49	-0.08		T-PLASMINOGEN ACTIVATOR F1-G; ITPG 7 CHAIN: NULL; ITPG 8	PLASMINOGEN ACTIVATION

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
435	1vap	A	70	184	1.8e-15	-0.14	0.00		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
435	1xka	L	70	154	5e-15	0.14	0.18		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
435	2not	A	34	137	5e-15	0.11	-0.13		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
435	9wga	A	20	181	1.7e-16	0.20	0.05		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
435	9wga	A	31	180	2.5e-27	0.32	-0.17		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
435	9wga	A	53	219	5e-30		79.05		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
435	9wga	A	55	229	3.4e-15	0.39	0.10		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
435	9wga	A	64	218	5e-30	0.79	-0.05		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
443	1ckl	A	12	132	8.5e-11	-0.11	0.05		CD46; CHAIN: A, B, C, D, E, F;	GLYCOPROTEIN MEMBRANE COFACTOR PROTEIN (MCP);

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										VIRUS RECEPTOR, COMPLEMENT COFACTOR, SHORT CONSENSUS REPEAT, 2 SCR, MEASLES VIRUS, GLYCOPROTEIN
443	1ckl	A	67	131	5e-11	0.36	0.03		CD46; CHAIN: A, B, C, D, E, F;	GLYCOPROTEIN MEMBRANE COFACTOR PROTEIN (MCP); VIRUS RECEPTOR, COMPLEMENT COFACTOR, SHORT CONSENSUS REPEAT, 2 SCR, MEASLES VIRUS, GLYCOPROTEIN
443	1e5g	A	72	192	3.4e-14	0.10	0.24		COMPLEMENT CONTROL PROTEIN; CHAIN: A;	COMPLEMENT INHIBITOR VCP, SP35; COMPLEMENT, NMR, MODULES, PROTEIN STRUCTURE, VACCINIA VIRUS
443	1e5g	A	73	155	1.3e-15	0.11	0.24		COMPLEMENT CONTROL PROTEIN; CHAIN: A;	COMPLEMENT INHIBITOR VCP, SP35; COMPLEMENT, NMR, MODULES, PROTEIN STRUCTURE, VACCINIA VIRUS
443	1hcc		73	132	5e-10	0.35	0.57		GLYCOPROTEIN 16TH COMPLEMENT CONTROL PROTEIN (/CCP$) OF FACTOR H 1HCC 3	
443	1hfh		70	192	1.3e-10			51.85	GLYCOPROTEIN FACTOR H, 15TH AND 16TH C-MODULE PAIR (NMR, MINIMIZED 1HFHA 1 AVERAGED STRUCTURE) 1HFH 4 1HFHA 5	
443	1hfh		71	155	1.3e-10	0.36	0.22		GLYCOPROTEIN FACTOR H,	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									15TH AND 16TH C-MODULE PAIR (NMR, MINIMIZED 1HFHA 1 AVERAGED STRUCTURE) 1HFH 4 1HFHA 5	
443	1qub	A	21	297	3.4e-27			56.46	HUMAN BETA2-GLYCOPROTEIN I; CHAIN: A;	MEMBRANE ADHESION SHORT CONSENSUS REPEAT, SUSHI, COMPLEMENT CONTROL PROTEIN, 2 N-GLYCOSYLATION, MULTI-DOMAIN, MEMBRANE ADHESION
443	1sfp		129	245	2.3e-27			51.76	ASFP; CHAIN: NULL;	SPERMADHESIN ACIDIC SEMINAL PROTEIN; SPERMADHESIN, BOVINE SEMINAL PLASMA PROTEIN, ACIDIC 2 SEMINAL FLUID PROTEIN, ASFP, CUB DOMAIN, X-RAY CRYSTAL 3 STRUCTURE, GROWTH FACTOR
443	1sfp		135	242	2.3e-27	0.38	0.81		ASFP; CHAIN: NULL;	SPERMADHESIN ACIDIC SEMINAL PROTEIN; SPERMADHESIN, BOVINE SEMINAL PLASMA PROTEIN, ACIDIC 2 SEMINAL FLUID PROTEIN, ASFP, CUB DOMAIN, X-RAY CRYSTAL 3 STRUCTURE, GROWTH FACTOR
443	1sfp		155	244	1.7e-07	0.01	0.16		ASFP; CHAIN: NULL;	SPERMADHESIN ACIDIC SEMINAL PROTEIN; SPERMADHESIN, BOVINE SEMINAL PLASMA PROTEIN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										ACIDIC 2 SEMINAL FLUID PROTEIN, ASFP, CUB DOMAIN, X-RAY CRYSTAL 3 STRUCTURE, GROWTH FACTOR
443	1spp	A	135	242	5e-28	0.38	0.71		MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-I; CHAIN: A; MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II; CHAIN: B	COMPLEX (SEMINAL PLASMA PROTEIN/SPP) SEMINAL PLASMA PROTEINS, SPERMADHESINS, CUB DOMAIN 2 ARCHITECTURE, COMPLEX (SEMINAL PLASMA PROTEIN/SPP)
443	1spp	B	129	242	5e-29	0.36	0.62		MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-I; CHAIN: A; MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II; CHAIN: B	COMPLEX (SEMINAL PLASMA PROTEIN/SPP) SEMINAL PLASMA PROTEINS, SPERMADHESINS, CUB DOMAIN 2 ARCHITECTURE, COMPLEX (SEMINAL PLASMA PROTEIN/SPP)
443	1vvc		10	127	3.4e-15	0.24	-0.14		VACCINIA VIRUS COMPLEMENT CONTROL PROTEIN; CHAIN: NULL;	COMPLEMENT INHIBITOR SP35, VCP, VACCINIA VIRUS SP35; COMPLEMENT INHIBITOR, COMPLEMENT MODULE, SCR, SUSHI DOMAIN, 2 MODULE PAIR
443	1vvc		72	194	1.7e-12	0.09	0.25		VACCINIA VIRUS COMPLEMENT CONTROL PROTEIN; CHAIN: NULL;	COMPLEMENT INHIBITOR SP35, VCP, VACCINIA VIRUS SP35; COMPLEMENT INHIBITOR, COMPLEMENT MODULE, SCR, SUSHI DOMAIN, 2 MODULE PAIR
443	1vvc		72	196	1.7e-12			50.78	VACCINIA VIRUS COMPLEMENT CONTROL	COMPLEMENT INHIBITOR SP35, VCP, VACCINIA VIRUS SP35;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									PROTEIN; CHAIN: NULL;	COMPLEMENT INHIBITOR, COMPLEMENT MODULE, SCR, SUSHI DOMAIN, 2 MODULE PAIR
454	1lba		230	375	5e-39	0.08	0.05		HYDROLASE(ACTING ON LINEAR AMIDES) LYSOZYME (E.C.3.5.1.28) MUTANT WITH ALA 6 REPLACED BY LYS 1LBA 3 AND RESIDUES 2 - 5 DELETED (DEL(2-5),A6K) 1LBA 4	
454	1lba		258	359	1.7e-23	0.03	0.33		HYDROLASE(ACTING ON LINEAR AMIDES) LYSOZYME (E.C.3.5.1.28) MUTANT WITH ALA 6 REPLACED BY LYS 1LBA 3 AND RESIDUES 2 - 5 DELETED (DEL(2-5),A6K) 1LBA 4	
454	1lba		72	232	2.5e-23			54.86	HYDROLASE(ACTING ON LINEAR AMIDES) LYSOZYME (E.C.3.5.1.28) MUTANT WITH ALA 6 REPLACED BY LYS 1LBA 3 AND RESIDUES 2 - 5 DELETED (DEL(2-5),A6K) 1LBA 4	
454	1lba		74	214	2.5e-23	0.26	0.55		HYDROLASE(ACTING ON LINEAR AMIDES) LYSOZYME (E.C.3.5.1.28) MUTANT WITH ALA 6 REPLACED BY LYS 1LBA 3 AND RESIDUES 2 - 5 DELETED (DEL(2-5),A6K)	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									1LBA 4	
454	1lba		81	175	3.4e-23	0.66	0.88		HYDROLASE(ACTING ON LINEAR AMIDES) LYSOZYME (E.C.3.5.1.28) MUTANT WITH ALA 6 REPLACED BY LYS 1LBA 3 AND RESIDUES 2 - 5 DELETED (DEL(2-5),A6K) 1LBA 4	
455	1c2a	A	35	148	0.0027	-0.45	0.03		BOWMAN-BIRK TRYPSIN INHIBITOR; CHAIN: A	HYDROLASE INHIBITOR ALL-BETA STRUCTURE, HYDROLASE INHIBITOR
458	1c17	M	110	248	1.2e-07			79.86	ATP SYNTHASE SUBUNIT C; CHAIN: A, B, C, D, E, F, G, H, I, J, K, L; ATP SYNTHASE SUBUNIT A; CHAIN: M;	MEMBRANE PROTEIN MEMBRANE PROTEIN, HELIX, COMPLEX
459	1fqv	A	28	67	0.005	-0.85	0.43		SKP2; CHAIN: A, C, E, G, I, K, M, O; SKP1; CHAIN: B, D, F, H, J, L, N, P;	LIGASE CYCLIN A/CDK2-ASSOCIATED PROTEIN P45; CYCLIN A/CDK2-ASSOCIATED PROTEIN P19; SKP1, SKP2, F-BOX, LRR, LEUCINE-RICH REPEAT, SCF, UBIQUITIN, 2 E3, UBIQUITIN PROTEIN LIGASE
474	1b8q	A	223	353	1.2e-13			50.34	NEURONAL NITRIC OXIDE SYNTHASE; CHAIN: A; HEPTAPEPTIDE; CHAIN: B;	OXIDOREDUCTASE PDZ DOMAIN, NNOS, NITRIC OXIDE SYNTHASE
474	1b8q	A	224	302	1.2e-13	0.05	0.88		NEURONAL NITRIC OXIDE SYNTHASE; CHAIN: A;	OXIDOREDUCTASE PDZ DOMAIN, NNOS, NITRIC OXIDE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									HEPTAPEPTIDE; CHAIN: B;	SYNTHASE
474	1b8q	A	313	429	1.8e-17	0.38	0.11		NEURONAL NITRIC OXIDE SYNTHASE; CHAIN: A;	OXIDOREDUCTASE PDZ DOMAIN, NNOS, NITRIC OXIDE SYNTHASE
474	1be9	A	116	229	1.7e-14	-0.22	0.70		HEPTAPEPTIDE; CHAIN: B; PSD-95; CHAIN: A; CRIPT; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
474	1be9	A	221	337	1.3e-09			51.39	PSD-95; CHAIN: A; CRIPT; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
474	1be9	A	230	338	1.3e-09	0.71	1.00		PSD-95; CHAIN: A; CRIPT; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
474	1be9	A	315	380	1e-10	0.07	0.18		PSD-95; CHAIN: A; CRIPT; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
474	1be9	A	349	413	1e-10	-0.39	0.28		PSD-95; CHAIN: A; CRIPT; CHAIN: B;	PEPTIDE RECOGNITION PEPTIDE RECOGNITION, PROTEIN LOCALIZATION
474	1ii6		231	330	2e-10	-0.19	0.01		INTERLEUKIN 16; CHAIN: NULL;	CYTOKINE LCF; CYTOKINE, LYMPHOCYTE CHEMOATTRACTANT FACTOR, PDZ DOMAIN
474	1ii6		282	413	2.5e-16			52.15	INTERLEUKIN 16; CHAIN: NULL;	CYTOKINE LCF; CYTOKINE, LYMPHOCYTE CHEMOATTRACTANT FACTOR, PDZ DOMAIN
474	1ii6		315	388	2.5e-16	0.93	1.00		INTERLEUKIN 16; CHAIN: NULL;	CYTOKINE LCF; CYTOKINE, LYMPHOCYTE CHEMOATTRACTANT FACTOR, PDZ DOMAIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
474	1kwa	A	234	321	1.5e-11	0.44	1.00		HCASK/LIN-2 PROTEIN; CHAIN: A, B;	KINASE HCASK, GLGF REPEAT, DHR; PDZ DOMAIN, NEUREXIN, SYNDECAN, RECEPTOR
474	1kwa	A	313	388	2.3e-16	0.21	1.00		HCASK/LIN-2 PROTEIN; CHAIN: A, B;	KINASE HCASK, GLGF REPEAT, DHR; PDZ DOMAIN, NEUREXIN, SYNDECAN, RECEPTOR CLUSTERING, KINASE
474	1pdr		122	218	1.2e-13	-0.25	0.27		HUMAN DISCS LARGE PROTEIN; CHAIN: NULL;	SIGNAL TRANSDUCTION HDLG, DHR3 DOMAIN; SIGNAL TRANSDUCTION, SH3 DOMAIN, REPEAT
474	1pdr		228	295	2.5e-12	0.23	0.99		HUMAN DISCS LARGE PROTEIN; CHAIN: NULL;	SIGNAL TRANSDUCTION HDLG, DHR3 DOMAIN; SIGNAL TRANSDUCTION, SH3 DOMAIN, REPEAT
474	1pdr		311	380	2e-12	0.41	0.82		HUMAN DISCS LARGE PROTEIN; CHAIN: NULL;	SIGNAL TRANSDUCTION HDLG, DHR3 DOMAIN; SIGNAL TRANSDUCTION, SH3 DOMAIN, REPEAT
474	1qau	A	117	224	7.5e-15	0.12	-0.01		NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: A;	OXIDOREDUCTASE BETA-FINGER
474	1qau	A	231	346	2.3e-13	0.45	0.83		NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: A;	OXIDOREDUCTASE BETA-FINGER
474	1qau	A	313	388	5e-16	0.88	1.00		NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: A;	OXIDOREDUCTASE BETA-FINGER
474	1qav	A	114	212	1.5e-15	0.01	1.00		ALPHA-1 SYNTROPHIN (RESIDUES 77-171); CHAIN: A;	MEMBRANE PROTEIN/OXIDOREDUCTASE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: B;	BETA-FINGER, HETERODIMER
474	1qav	A	229	309	7.5e-12	0.68	1.00		ALPHA-1 SYNTROPHIN (RESIDUES 77-171); CHAIN: A; NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: B;	MEMBRANE PROTEIN/OXIDOREDUCTASE BETA-FINGER, HETERODIMER
474	1qav	A	311	388	2e-16	0.66	1.00		ALPHA-1 SYNTROPHIN (RESIDUES 77-171); CHAIN: A; NEURONAL NITRIC OXIDE SYNTHASE (RESIDUES 1-130); CHAIN: B;	MEMBRANE PROTEIN/OXIDOREDUCTASE BETA-FINGER, HETERODIMER
474	1qlc	A	116	213	5e-15	0.81	0.89		POSTSYNAPTIC DENSITY PROTEIN 95; CHAIN: A;	PEPTIDE RECOGNITION PSD-95; PDZ DOMAIN, NEURONAL NITRIC OXIDE SYNTHASE, NMDA RECEPTOR 2 BINDING
474	1qlc	A	120	213	5.1e-15	0.36	0.22		POSTSYNAPTIC DENSITY PROTEIN 95; CHAIN: A;	PEPTIDE RECOGNITION PSD-95; PDZ DOMAIN, NEURONAL NITRIC OXIDE SYNTHASE, NMDA RECEPTOR 2 BINDING
474	1qlc	A	229	309	1.5e-09	0.68	1.00		POSTSYNAPTIC DENSITY PROTEIN 95; CHAIN: A;	PEPTIDE RECOGNITION PSD-95; PDZ DOMAIN, NEURONAL NITRIC OXIDE SYNTHASE, NMDA RECEPTOR 2 BINDING
474	1qlc	A	311	388	1.5e-14	0.75	1.00		POSTSYNAPTIC DENSITY PROTEIN 95; CHAIN: A;	PEPTIDE RECOGNITION PSD-95; PDZ DOMAIN, NEURONAL NITRIC OXIDE SYNTHASE, NMDA RECEPTOR 2 BINDING
474	3pdz	A	113	212	5e-15	0.36	0.96		TYROSINE PHOSPHATASE (PTP-BAS, TYPE 1); CHAIN: A;	HYDROLASE PDZ DOMAIN, HUMAN PHOSPHATASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										HPTP1E, PTP-BAS, SPECIFICITY 2 OF BINDING
474	3pdz	A	227	324	7.5e-12	0.47	0.99		TYROSINE PHOSPHATASE (PTP-BAS, TYPE 1); CHAIN: A;	HYDROLASE PDZ DOMAIN, HUMAN PHOSPHATASE, HPTP1E, PTP-BAS, SPECIFICITY 2 OF BINDING
474	3pdz	A	311	388	2.5e-15	1.23	1.00		TYROSINE PHOSPHATASE (PTP-BAS, TYPE 1); CHAIN: A;	HYDROLASE PDZ DOMAIN, HUMAN PHOSPHATASE, HPTP1E, PTP-BAS, SPECIFICITY 2 OF BINDING
477	1a0j	A	36	265	0			235.32	TRYPSIN; CHAIN: A, B, C, D;	SERINE PROTEASE SERINE PROTEINASE, TRYPSIN, HYDROLASE
477	1a0j	A	36	265	0	1.07	1.00		TRYPSIN; CHAIN: A, B, C, D;	SERINE PROTEASE SERINE PROTEINASE, TRYPSIN, HYDROLASE
477	1a0l	A	36	264	5.1e-82			167.61	BETA-TRYPTASE; CHAIN: A, B, C, D;	SERINE PROTEINASE TRYPSIN-LIKE SERINE PROTEINASE, TETRAMER, HEPARIN, ALLERGY, 2 ASTHMA
477	1a5i	A	23	263	2.5e-83			173.85	PLASMINOGEN ACTIVATOR; CHAIN: A; GLU-GLY-ARG CHLOROMETHYL KETONE; CHAIN: I;	COMPLEX (SERINE PROTEASE/INHIBITOR) (DELTAFEK)DSPAALPHA1; EGRCMK; SERINE PROTEASE, FIBRINOLYTIC ENZYMES, PLASMINOGEN 2 ACTIVATORS
477	1ao5	A	36	266	2.5e-96			226.34	GLANDULAR KALLIKREIN-13; CHAIN: A, B;	SERINE PROTEASE PRORENIN CONVERTING ENZYME (PRECE), EPIDERMAL GLANDULAR KALLIKREIN,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										SERINE PROTEASE, PROTEIN MATURATION
477	1ao5	A	38	264	2.5e-96	1.16	1.00		GLANDULAR KALLIKREIN-13; CHAIN: A, B;	SERINE PROTEASE PRORENIN CONVERTING ENZYME (PRECE), EPIDERMAL GLANDULAR KALLIKREIN, SERINE PROTEASE, PROTEIN MATURATION
477	1aut	C	36	263	2.2e-88			172.05	ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MA; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
477	1bio		36	263	5e-89			198.56	COMPLEMENT FACTOR D; CHAIN: NULL;	SERINE PROTEASE SERINE PROTEASE, HYDROLASE, COMPLEMENT, FACTOR D, CATALYTIC 2 TRIAD, SELF-REGULATION
477	1bqy	A	36	271	1e-92			205.81	PLASMINOGEN ACTIVATOR; CHAIN: A, B; GLU-GLY-ARG-CHLOROMETHYLKETONE INHIBITOR; CHAIN: E, F;	BLOOD CLOTTING TSV-PA; FIBRINOLYSIS, PLASMINOGEN ACTIVATOR, SERINE PROTEINASE, 2 SNAKE VENOM, COMPLEX (HYDROLASE/INHIBITOR), BLOOD CLOTTING
477	1cgh	A	36	264	3.4e-74			175.67	CATHEPSIN G; CHAIN: A; PHOSPHONATE INHIBITOR SUC-VAL-PRO-PHEP-(OPH)2;	COMPLEX (SERINE PROTEASE/INHIBITOR) INFLAMMATION, INHIBITOR,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: S;	SPECIFICITY, SERINE PROTEASE, 2 COMPLEX (SERINE PROTEASE/INHIBITOR)
477	1dpo		36	265	1e-97			226.55	TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, ZYMOGEN, 2 SIGNAL, MULTIGENE FAMILY
477	1fxy	A	36	266	1.7e-91			218.47	COAGULATION FACTOR XA-TRYPSIN CHIMERA; CHAIN: A; D-PHE-PRO-ARG-CHLOROMETHYLKETONE (PPACK) WITH CHAIN: I;	COMPLEX (PROTEASE/INHIBITOR) TRYPSIN, COAGULATION FACTOR XA, CHIMERA, PROTEASE, PPACK, 2 CHLOROMETHYLKETONE, COMPLEX (PROTEASE/INHIBITOR)
477	1mct	A	36	265	0			234.27	COMPLEX(PROTEINASE/INHIBITOR) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH INHIBITOR FROM BITTER 1MCT 3 GOURD 1MCT 4	
477	1mct	A	36	265	0	1.20	1.00		COMPLEX(PROTEINASE/INHIBITOR) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH INHIBITOR FROM BITTER 1MCT 3 GOURD 1MCT 4	
477	1npm	A	36	263	5e-94			253.14	NEUROPSIN; CHAIN: A, B;	SERINE PROTEINASE SERINE PROTEINASE, GLYCOPROTEIN
477	1pfx	C	36	263	5e-91			177.42	FACTOR IXA; CHAIN: C, L;; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
477	1qrz	A	21	265	1.7e-88			176.03	PLASMINOGEN; CHAIN: A, B, C, D;	HYDROLASE MICROPLASMINOGEN, SERINE PROTEASE, ZYMOGEN, CHYMOTRYPSIN 2 FAMILY, HYDROLASE
477	1rfn	A	36	263	1.3e-90			177.83	COAGULATION FACTOR IX; CHAIN: A; COAGULATION FACTOR IX; CHAIN: B;	COAGULATION FACTOR SERINE PROTEINASE, BLOOD COAGULATION, COAGULATION FACTOR
477	1rtf	B	36	264	1e-84			177.39	TWO CHAIN TISSUE PLASMINOGEN ACTIVATOR; CHAIN: A, B;	SERINE PROTEASE (TC)-T-PA; SERINE PROTEASE, FIBRINOLYTIC ENZYMES
477	1sgf	A	45	266	7.5e-83			179.65	NERVE GROWTH FACTOR; CHAIN: A, B, G, X, Y, Z;	GROWTH FACTOR 7S NGF; GROWTH FACTOR (BETA-NGF), HYDROLASE - SERINE PROTEINASE 2 (GAMMA-NGF), INACTIVE SERINE PROTEINASE (ALPHA-NGF)
477	1sgf	G	36	265	8.5e-99	1.03	1.00		NERVE GROWTH FACTOR; CHAIN: A, B, G, X, Y, Z;	GROWTH FACTOR 7S NGF; GROWTH FACTOR (BETA-NGF), HYDROLASE - SERINE PROTEINASE 2 (GAMMA-NGF), INACTIVE SERINE PROTEINASE (ALPHA-NGF)
477	1sgf	G	36	266	8.5e-99			248.29	NERVE GROWTH FACTOR;	GROWTH FACTOR 7S NGF;

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, B, G, X, Y, Z;	GROWTH FACTOR (BETA-NGF), HYDROLASE - SERINE PROTEINASE 2 (GAMMA-NGF), INACTIVE SERINE PROTEINASE (ALPHA-NGF)
477	1slw	B	36	265	1e-99			223.23	ECOTIN; CHAIN: A; ANIONIC TRYPSIN; CHAIN: B;	COMPLEX (SERINE PROTEASE/INHIBITOR) TRYPSIN INHIBITOR; SERINE PROTEASE, INHIBITOR, COMPLEX, METAL BINDING SITES, 2 PROTEIN ENGINEERING, PROTEASE-SUBSTRATE INTERACTIONS, 3 METALLOPROTEINS
477	1slw	B	36	265	1e-99	1.25	1.00		ECOTIN; CHAIN: A; ANIONIC TRYPSIN; CHAIN: B;	COMPLEX (SERINE PROTEASE/INHIBITOR) TRYPSIN INHIBITOR; SERINE PROTEASE, INHIBITOR, COMPLEX, METAL BINDING SITES, 2 PROTEIN ENGINEERING, PROTEASE-SUBSTRATE INTERACTIONS, 3 METALLOPROTEINS
477	1ton		36	266	5e-97			232.87	HYDROLASE(SERINE PROTEINASE) TONIN (E.C. NUMBER NOT ASSIGNED) ITON 4	
477	1ton		38	264	5e-97	1.16	1.00		HYDROLASE(SERINE PROTEINASE) TONIN (E.C. NUMBER NOT ASSIGNED) ITON 4	
477	1tm	A	36	266	0			229.81	HYDROLASE (SERINE	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH THE INHIBITOR 1TRN 3 DIISOPROPYL-FLUOROPHOSPHOFLUORIDATE (DFP) 1TRN 4 HUMAN TRYPSIN, DFP INHIBITED 1TRN 6	
477	1trn	A	36	266	0	1.00	1.00		HYDROLASE (SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH THE INHIBITOR 1TRN 3 DIISOPROPYL-FLUOROPHOSPHOFLUORIDATE (DFP) 1TRN 4 HUMAN TRYPSIN, DFP INHIBITED 1TRN 6	
477	2tbs		36	265	1e-99			222.15	HYDROLASE(SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH BENZAMIDINE INHIBITOR 2TBS 3	
477	2tbs		36	265	1e-99	1.04	1.00		HYDROLASE(SERINE PROTEINASE) TRYPSIN (E.C.3.4.21.4) COMPLEXED WITH BENZAMIDINE INHIBITOR 2TBS 3	
477	5ptp		36	265	0			230.52	BETA TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, 2 ZYMOGEN, SIGNAL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
477	5ptp		36	265	0	1.26	1.00		BETA TRYPSIN; CHAIN: NULL;	SERINE PROTEASE HYDROLASE, SERINE PROTEASE, DIGESTION, PANCREAS, 2 ZYMOGEN, SIGNAL
487	1epf	A	18	124	5e-07	0.22	0.33		NEURAL CELL ADHESION MOLECULE; CHAIN: A, B, C, D;	CELL ADHESION NCAM; NCAM, IMMUNOGLOBULIN FOLD, GLYCOPROTEIN
487	1f5w	A	15	107	1e-06	0.02	0.36		COXSACKIE VIRUS AND ADENOVIRUS RECEPTOR; CHAIN: A, B;	VIRUS/VIRAL PROTEIN RECEPTOR IMMUNOGLOBULIN V DOMAIN FOLD, SYMMETRIC DIMER
487	1fhg	A	22	109	2.5e-07	0.09	0.06		TELOKIN; CHAIN: A	CONTRACTILE PROTEIN IMMUNOGLOBULIN FOLD, BETA BARREL
487	1tnm		22	107	2.5e-06	0.11	0.11		MUSCLE PROTEIN TITIN MODULE M5 (CONNECTIN) ITNM 3 (NMR, MINIMIZED AVERAGE STRUCTURE) ITNM 4 ITNM 58	
487	2ncm		20	109	2e-06	0.12	0.30		NEURAL CELL ADHESION MOLECULE; CHAIN: NULL;	CELL ADHESION NCAM DOMAIN 1; CELL ADHESION, GLYCOPROTEIN, HEPARIN-BINDING, GPI-ANCHOR, 2 NEURAL ADHESION MOLECULE, IMMUNOGLOBULIN FOLD, SIGNAL
487	3ncm	A	22	109	5e-07	-0.09	0.12		NEURAL CELL ADHESION MOLECULE, LARGE	CELL ADHESION PROTEIN NCAM MODULE 2; CELL

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									ISOFORM; CHAIN: A;	ADHESION, GLYCOPROTEIN, HEPARIN-BINDING, GPI-ANCHOR, 2 NEURAL ADHESION MOLECULE, IMMUNOGLOBULIN FOLD, HOMOPHILIC 3 BINDING, CELL ADHESION PROTEIN
492	1sfp		38	78	0.0015	-0.73	0.71		ASFP; CHAIN: NULL;	SPERMADHESIN ACIDIC SEMINAL PROTEIN; SPERMADHESIN, BOVINE SEMINAL PLASMA PROTEIN, ACIDIC 2 SEMINAL FLUID PROTEIN, ASFP, CUB DOMAIN, X-RAY CRYSTAL 3 STRUCTURE, GROWTH FACTOR
492	1spp	B	24	78	0.0015	-0.08	0.10		MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-I; CHAIN: A; MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II; CHAIN: B	COMPLEX (SEMINAL PLASMA PROTEIN/SPP) SEMINAL PLASMA PROTEINS, SPERMADHESINS, CUB DOMAIN 2 ARCHITECTURE, COMPLEX (SEMINAL PLASMA PROTEIN/SPP)
496	1cfe		60	219	5.1e-42	0.22	1.00		PATHOGENESIS-RELATED PROTEIN P14A; CHAIN: NULL;	PATHOGENESIS-RELATED PROTEIN PATHOGENESIS-RELATED LEAF PROTEIN 6, ETHYLENE PATHOGENESIS-RELATED PROTEIN, PR-1 PROTEINS, 2 PLANT DEFENSE
496	1cfe		61	219	5.1e-42			75.76	PATHOGENESIS-RELATED	PATHOGENESIS-RELATED

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									PROTEIN P14A; CHAIN: NULL;	PROTEIN PATHOGENESIS-RELATED LEAF PROTEIN 6, ETHYLENE PATHOGENESIS-RELATED PROTEIN, PR-1 PROTEINS, 2 PLANT DEFENSE
496	1qnx	A	58	220	1.7e-42	0.33	1.00		VES V 5; CHAIN: A;	ALLERGEN ANTIGEN 5; ANTIGEN 5, ALLERGEN, VESPID VENOM
511	1def		63	229	3.4e-46			62.32	PEPTIDE DEFORMYLASE; CHAIN: NULL;	HYDROLASE HYDROLASE, ZINC METALLOPROTEASE
512	1c44	A	92	211	8.5e-37	0.82	0.99		STEROL CARRIER PROTEIN 2; CHAIN: A;	LIPID BINDING PROTEIN NON SPECIFIC LIPID BINDING PROTEIN; STEROL CARRIER PROTEIN, NON SPECIFIC LIPID TRANSFER PROTEIN, 2 FATTY ACID BINDING, FATTY ACYL COA BINDING
513	1edh	A	46	166	3.4e-17	-0.18	0.25		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING PROTEIN
513	1edh	A	51	164	1.7e-23	0.24	0.98		E-CADHERIN; CHAIN: A, B;	CELL ADHESION PROTEIN EPITHELIAL CADHERIN DOMAINS 1 AND 2, ECAD12; CADHERIN, CELL ADHESION PROTEIN, CALCIUM BINDING

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										PROTEIN
513	1ncg		45	138	7.5e-17	0.44	0.93		N-CADHERIN; INCG 3	CELL ADHESION PROTEIN CADHERIN INCG 13
513	1nci	B	45	140	5e-16	-0.22	0.63		N-CADHERIN; INCI 3	CELL ADHESION PROTEIN CADHERIN INCI 13
513	1ncj	A	45	164	5e-20	0.22	0.86		N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
513	1ncj	A	45	165	3.4e-19	-0.11	0.36		N-CADHERIN; CHAIN: A;	CELL ADHESION PROTEIN CELL ADHESION PROTEIN
513	1suh		44	144	5e-24			55.29	EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
513	1suh		45	144	3.4e-09	0.31	0.59		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
513	1suh		45	144	5e-24	0.66	0.98		EPITHELIAL CADHERIN; CHAIN: NULL;	CELL ADHESION UVOMORULIN; CADHERIN, CALCIUM BINDING, CELL ADHESION
516	1hcn	B	15	126	1e-45			94.05	HORMONE HUMAN CHORIONIC GONADOTROPIN 1HCN 3	
516	1hcn	B	16	125	1e-45	0.10	1.00		HORMONE HUMAN CHORIONIC GONADOTROPIN 1HCN 3	
516	1hcn	B	17	126	1.7e-43	-0.14	1.00		HORMONE HUMAN CHORIONIC GONADOTROPIN 1HCN 3	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
544	1a14	H	20	139	5.1e-14			61.90	NEURAMINIDASE; CHAIN: N; SINGLE CHAIN ANTIBODY; CHAIN: H, L;	COMPLEX (ANTIBODY/ANTIGEN) COMPLEX (ANTIBODY/ANTIGEN), SINGLE-CHAIN ANTIBODY, 2 GLYCOSYLATED PROTEIN
544	1a2y	A	19	134	3.4e-28			57.90	MONOCLONAL ANTIBODY D1.3; CHAIN: A, B; LYSOZYME; CHAIN: C;	COMPLEX (IMMUNOGLOBULIN/HYDROLASE) COMPLEX (IMMUNOGLOBULIN/HYDROLASE), IMMUNOGLOBULIN V 2 REGION, SIGNAL, HYDROLASE, GLYCOSIDASE, BACTERIOLYTIC 3 ENZYME, EGG WHITE
544	1a7q	L	19	132	3.4e-26			56.95	MONOCLONAL ANTIBODY D1.3; CHAIN: L, H;	IMMUNOGLOBULIN IMMUNOGLOBULIN, VARIANT
544	1adq	L	21	141	6.8e-46	0.39	0.95		IGG4 REA; CHAIN: A; RF-AN IGM/LAMBDA; CHAIN: H, L;	COMPLEX (IMMUNOGLOBULIN/AUTOANTIGEN) COMPLEX (IMMUNOGLOBULIN/AUTOANTIGEN), RHEUMATOID FACTOR 2 AUTO-ANTIBODY COMPLEX
544	1ao7	D	20	142	2.3e-19			59.00	HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA-A2 HEAVY CHAIN; CLASS I MHC, T-CELL RECEPTOR, VIRAL PEPTIDE, 2 COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
544	1ap2	A	19	133	3.4e-31			57.34	MONOCLONAL ANTIBODY C219; CHAIN: A, B, C, D;	IMMUNOGLOBULIN VARIABLE DOMAIN; SINGLE CHAIN FV, MONOCLONAL ANTIBODY, C219, P-GLYCOPROTEIN, 2
544	1aqk	L	22	141	3.4e-50	0.07	0.95		FAB B7-15A2; CHAIN: L, H;	IMMUNOGLOBULIN HUMAN FAB, ANTI-TETANUS TOXOID, HIGH AFFINITY, CRYSTAL 2 PACKING MOTIF, PROGRAMMING PROPENSITY TO CRYSTALLIZE, 3 IMMUNOGLOBULIN
544	1ar1	D	19	129	1e-24			57.17	CYTOCHROME C OXIDASE; CHAIN: A, B; ANTIBODY FV FRAGMENT; CHAIN: C, D;	COMPLEX (OXIDOREDUCTASE/ANTIBODY) CYTOCHROME AA3, COMPLEX IV, FERROCYTOCHROME C, COMPLEX (OXIDOREDUCTASE/ANTIBODY), ELECTRON TRANSPORT, 2 TRANSMEMBRANE, CYTOCHROME OXIDASE, ANTIBODY COMPLEX
544	1b0w	A	19	127	1.7e-29			58.09	BENCE-JONES KAPPA I PROTEIN BRE; CHAIN: A, B, C;	IMMUNE SYSTEM BENCE-JONES; IMMUNOGLOBULIN, AMYLOID, IMMUNE SYSTEM
544	1bfv	L	19	135	8.5e-28			57.62	FV4155; CHAIN: L, H;	IMMUNOGLOBULIN IMMUNOGLOBULIN, FV FRAGMENT, STEROID HORMONE, 2 FINE SPECIFICITY
544	1bjm	A	21	142	5.1e-45	0.33	0.90		LOC - LAMBDA I TYPE LIGHT-CHAIN DIMER; IBJM 6 CHAIN: A, B; IBJM 7	IMMUNOGLOBULIN BENCE-JONES PROTEIN; IBJM 8 BENCE JONES, ANTIBODY, MULTIPLE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										QUATERNARY STRUCTURES IBJM 13
544	1bvk	A	19	135	5.1e-32			61.61	HULYS11; CHAIN: A, B, D, E; LYSOZYME; CHAIN: C, F;	COMPLEX (HUMANIZED ANTIBODY/HYDROLASE) MURAMIDASE; HUMANIZED ANTIBODY, ANTIBODY COMPLEX, FV, ANTI-LYSOZYME, 2 COMPLEX (HUMANIZED ANTIBODY/HYDROLASE)
544	1bww	A	17	132	1.7e-31			61.59	IG KAPPA CHAIN V-I REGION REI; CHAIN: A, B;	IMMUNE SYSTEM REIV, STABILIZED IMMUNOGLOBULIN FRAGMENT, BENCE-JONES 2 PROTEIN, IMMUNE SYSTEM
544	1cd0	A	22	122	1.2e-46	0.57	1.00		JTO, A VARIABLE DOMAIN FROM LAMBDA-6 TYPE CHAIN: A, B;	IMMUNE SYSTEM IMMUNOGLOBULIN, BENCE-JONES PROTEIN, LAMDA-6
544	1dlf	L	19	135	3.4e-27			57.84	ANTI-DANSYL IMMUNOGLOBULIN IGG2A(S); CHAIN: L, H;	IMMUNOGLOBULIN ANTI-DANSYL FV FRAGMENT FV FRAGMENT, IMMUNOGLOBULIN
544	1fgv	L	19	134	1.7e-33			67.61	IMMUNOGLOBULIN FV FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 1FGV 3 ANTIBODY 'H52' (HUH52-AA FV) 1FGV 4	
544	1fvc	A	19	136	3.4e-31			64.02	IMMUNOGLOBULIN FV FRAGMENT OF HUMANIZED ANTIBODY 4D5, VERSION 8	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									1FVC 3	
544	1igm	L	19	144	1e-30			62.66	IMMUNOGLOBULIN IMMUNOGLOBULIN M (IG-M) FV FRAGMENT 1IGM 3	
544	1maj		19	135	1.2e-25			56.76	IMMUNOGLOBULIN MURINE ANTIBODY 26-10 VL DOMAIN (NMR, 15 ENERGY MINIMIZED 1MAJ 3 STRUCTURES) 1MAJ 4	
544	1mel	A	22	145	3.4e-12			58.28	VH SINGLE-DOMAIN ANTIBODY; CHAIN: A, B; LYSOZYME; CHAIN: L, M;	COMPLEX (ANTIBODY/ANTIGEN) CAB-LYS3 COMPLEX; CAMEL SINGLE-DOMAIN ANTI-LYSOZYME, COMPLEX 2 (ANTIBODY/ANTIGEN)
544	1rvf	L	20	138	1.2e-31			62.39	HUMAN RHINOVIRUS 14 COAT PROTEIN; CHAIN: 1, 2, 3, 4; FAB 17-1A; CHAIN: L, H	COMPLEX (COAT PROTEIN/IMMUNOGLOBULIN) POLYPROTEIN, COAT PROTEIN, CORE PROTEIN, RNA-DIRECTED RNA 2 POLYMERASE, HYDROLASE, THIOL PROTEASE, MYRISTYLATION, 3 COMPLEX (COAT PROTEIN/IMMUNOGLOBULIN)
544	1wtl	A	19	127	1.5e-31			60.42	IMMUNOGLOBULIN WAT, A VARIABLE DOMAIN FROM IMMUNOGLOBULIN LIGHT-CHAIN 1WTL 3 (BENCE-JONES PROTEIN) 1WTL 4	
544	2cd0	A	23	122	5.1e-47	0.67	1.00		BENCE-JONES PROTEIN WIL, A VARIABLE DOMAIN FROM	IMMUNE SYSTEM IMMUNOGLOBULIN, BENCE-

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN: A, B;	JONES PROTEIN, LAMBDA-6
544	2fb4	L	20	142	1.7e-46	0.21	0.86		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB 2FB4 4	
544	2imn		19	127	6.8e-33			60.82	IMUNOGLOBULIN IMMUNOGLOBULIN VL DOMAIN (VARIABLE DOMAIN OF KAPPA 2IMN 3 LIGHT CHAIN) OF MCPC603 MUTANT IN WHICH 2IMN 4 COMPLEMENTARITY-DETERMINING REGION I HAS BEEN REPLACED BY 2IMN 5 THAT FROM MOPC167 2IMN 6	
544	2mcg	1	21	142	1.7e-52	0.32	0.72		IMMUNOGLOBULIN IMMUNOGLOBULIN LAMBDA LIGHT CHAIN DIMER (/MCGS) 2MCG 3 (TRIGONAL FORM) 2MCG 4	
544	2rhe		20	140	1.2e-44			68.90	IMMUNOGLOBULIN BENCE-*JONES PROTEIN (LAMBDA, VARIABLE DOMAIN) 2RHE 4	
544	2rhe		21	121	1.2e-44	0.59	1.00		IMMUNOGLOBULIN BENCE-*JONES PROTEIN (LAMBDA, VARIABLE DOMAIN) 2RHE 4	
544	43c9	A	19	134	1.7e-31			57.86	IMMUNOGLOBULIN (LIGHT CHAIN); CHAIN: A, C, E, G; IMMUNOGLOBULIN (HEAVY CHAIN); CHAIN: B, D, F, H;	IMMUNOGLOBULIN IMMUNOGLOBULIN
544	43c9	B	18	140	5.1e-15			61.03	IMMUNOGLOBULIN (LIGHT CHAIN); CHAIN: A, C, E, G;	IMMUNOGLOBULIN

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									IMMUNOGLOBULIN (HEAVY CHAIN); CHAIN: B, D, F, H;	
544	7fab	L	21	141	1.4e-43	0.29	0.21		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB' NEW (LAMBDA LIGHT CHAIN) 7FAB 3	
544	8fab	A	23	141	3.4e-44	0.31	0.84		IMMUNOGLOBULIN FAB FRAGMENT FROM HUMAN IMMUNOGLOBULIN IGG1 (LAMBDA, HIL) 8FAB 3	
548	1fxx	A	224	454	6.8e-97	0.07	1.00		EXONUCLEASE I; CHAIN: A;	HYDROLASE EXODEOXYRIBONUCLEASE I; ALPHA-BETA DOMAIN, SH3-LIKE DOMAIN, DNAQ SUPERFAMILY
557	1ao7	E	74	194	3.4e-53	0.46	1.00		HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA-A2 HEAVY CHAIN; CLASS I MHC, T-CELL RECEPTOR, VIRAL PEPTIDE, 2 COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR
557	1ao7	E	74	217	3.4e-53			80.28	HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA-A2 HEAVY CHAIN; CLASS I MHC, T-CELL RECEPTOR, VIRAL PEPTIDE, 2 COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
557	1bd2	E	74	194	1e-55	0.58	1.00		HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA A2 HEAVY CHAIN; COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR)
557	1bd2	E	74	217	1e-55			62.05	HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA A2 HEAVY CHAIN; COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR)
557	1bec		74	217	8.5e-57			72.58	14.3.D T CELL ANTIGEN RECEPTOR; IBEC 5 CHAIN: NULL; IBEC 6	RECEPTOR T CELL RECEPTOR IBEC 14
557	1bec		75	195	8.5e-57	0.51	1.00		14.3.D T CELL ANTIGEN RECEPTOR; IBEC 5 CHAIN: NULL; IBEC 6	RECEPTOR T CELL RECEPTOR IBEC 14
557	1bwm	A	74	217	1.7e-48			61.36	ALPHA-BETA T CELL RECEPTOR (TCR) (D10); CHAIN: A;	IMMUNE SYSTEM IMMUNOGLOBULIN, IMMUNORECEPTOR, IMMUNE SYSTEM
557	1bwm	A	75	185	1.7e-48	0.55	1.00		ALPHA-BETA T CELL RECEPTOR (TCR) (D10); CHAIN: A;	IMMUNE SYSTEM IMMUNOGLOBULIN, IMMUNORECEPTOR, IMMUNE SYSTEM
557	1d9k	B	75	185	1.7e-48	0.63	1.00		T-CELL RECEPTOR D10 (ALPHA CHAIN); CHAIN: A, E; T-CELL RECEPTOR D10 (BETA CHAIN); CHAIN: B, F; MHC I-AK A CHAIN (ALPHA	IMMUNE SYSTEM MHC I-AK; MHC I-AK; T-CELL RECEPTOR, MHC CLASS II, D10, I-AK

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHAIN); CHAIN: C, G; MHC I-AK B CHAIN (BETA CHAIN); CHAIN: D, H; CONALBUMIN PEPTIDE; CHAIN: P, Q;	
557	1fyt	E	74	194	3.4e-50	0.39	1.00		HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR CHAIN: A; HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR-1 CHAIN: B; HEMAGGLUTININ HA1 PEPTIDE CHAIN; CHAIN: C; T-CELL RECEPTOR ALPHA CHAIN; CHAIN: D; T-CELL RECEPTOR BETA CHAIN; CHAIN: E;	IMMUNE SYSTEM HLA-DR1, DRA; HLA-DR1, DRB1 0101; TCR HA1.7 ALPHA CHAIN; TCR HA1.7 BETA CHAIN; PROTEIN-PROTEIN COMPLEX, IMMUNOGLOBULIN FOLD
557	1nct		27	93	0.0015	-0.00	0.11		TITIN; CHAIN: NULL;	MUSCLE PROTEIN CONNECTIN, NEXTM5; CELL ADHESION, GLYCOPROTEIN, TRANSMEMBRANE, REPEAT, BRAIN, 2 IMMUNOGLOBULIN FOLD, ALTERNATIVE SPLICING, SIGNAL, 3 MUSCLE PROTEIN
557	1tcr	B	72	195	8.5e-55	0.45	1.00		ALPHA, BETA T-CELL RECEPTOR CHAIN: A, B;	RECEPTOR TCR, T-CELL, RECEPTOR, TRANSMEMBRANE, GLYCOPROTEIN, SIGNAL
557	1tcr	B	72	217	8.5e-55			74.81	ALPHA, BETA T-CELL RECEPTOR CHAIN: A, B;	RECEPTOR TCR, T-CELL, RECEPTOR, TRANSMEMBRANE, GLYCOPROTEIN, SIGNAL
559	1evs	A	29	212	1.8e-78	1.02	1.00		ONCOSTATIN M; CHAIN: A;	CYTOKINE 4-HELIX BUNDLE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										GP130 BINDING CYTOKINE
559	1evs	A	29	212	5.1e-76	1.13	1.00		ONCOSTATIN M; CHAIN: A;	CYTOKINE 4-HELIX BUNDLE, GP130 BINDING CYTOKINE
568	1a4y	A	52	166	5e-05	0.05	0.43		RIBONUCLEASE INHIBITOR; CHAIN: A, D; ANGIOGENIN; CHAIN: B, E;	COMPLEX (INHIBITOR/NUCLEASE) COMPLEX (INHIBITOR/NUCLEASE), COMPLEX (RI-ANG), HYDROLASE 2 MOLECULAR RECOGNITION, EPITOPE MAPPING, LEUCINE-RICH 3 REPEATS
592	1ac6	A	27	119	1.3e-15			56.47	T-CELL RECEPTOR ALPHA; CHAIN: A, B;	RECEPTOR RECEPTOR, V ALPHA DOMAIN, SITE-DIRECTED MUTAGENESIS, 2 THREE-DIMENSIONAL STRUCTURE, GLYCOPROTEIN, SIGNAL
592	1ao7	D	26	119	7.5e-21			51.88	HLA-A 0201; CHAIN: A; BETA-2 MICROGLOBULIN; CHAIN: B; TAX PEPTIDE; CHAIN: C; T CELL RECEPTOR ALPHA; CHAIN: D; T CELL RECEPTOR BETA; CHAIN: E;	COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR) HLA-A2 HEAVY CHAIN; CLASS I MHC, T-CELL RECEPTOR, VIRAL PEPTIDE, 2 COMPLEX (MHC/VIRAL PEPTIDE/RECEPTOR
592	1aqk	L	28	117	5.1e-48	0.35	0.89		FAB B7-15A2; CHAIN: L, H;	IMMUNOGLOBULIN HUMAN FAB, ANTI-TETANUS TOXOID, HIGH AFFINITY, CRYSTAL 2 PACKING MOTIF,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										PROGRAMMING PROPENSITY TO CRYSTALLIZE, 3 IMMUNOGLOBULIN
592	1b6d	A	25	114	1.2e-44	0.16	0.60		IMMUNOGLOBULIN; CHAIN: A, B;	IMMUNOGLOBULIN IMMUNOGLOBULIN, KAPPA LIGHT-CHAIN DIMER HEADER
592	1bj1	L	25	114	5.1e-46	0.37	0.63		FAB FRAGMENT; CHAIN: L, H, J, K; VASCULAR ENDOTHELIAL GROWTH FACTOR; CHAIN: V, W;	COMPLEX (ANTIBODY/ANTIGEN) FAB-12; VEGF; COMPLEX (ANTIBODY/ANTIGEN), ANGIOGENIC FACTOR
592	1bjm	A	27	116	5.1e-45	0.13	0.83		LOC - LAMBDA I TYPE LIGHT-CHAIN DIMER; 1BJM 6 CHAIN: A, B; 1BJM 7	IMMUNOGLOBULIN BENCE-JONES PROTEIN; 1BJM 8 BENCE JONES, ANTIBODY, MULTIPLE QUATERNARY STRUCTURES 1BJM 13
592	1bww	A	23	114	1.7e-45	0.27	0.31		IG KAPPA CHAIN V-I REGION REI; CHAIN: A, B;	IMMUNE SYSTEM REIV, STABILIZED IMMUNOGLOBULIN FRAGMENT, BENCE-JONES 2 PROTEIN, IMMUNE SYSTEM
592	1dee	A	25	114	3.4e-47	0.25	0.48		IGM RF 2A2; CHAIN: A, C, E; IGM RF 2A2; CHAIN: B, D, F; IMMUNOGLOBULIN G BINDING PROTEIN A; CHAIN: G, H;	IMMUNE SYSTEM FAB-IBP COMPLEX CRYSTAL STRUCTURE 2.7A RESOLUTION BINDING 2 OUTSIDE THE ANTIGEN COMBINING SITE SUPERANTIGEN FAB VH3 3 SPECIFICITY
592	1dfb	L	25	119	8.5e-47	0.50	0.64		IMMUNOGLOBULIN 3D6 FAB 1DFB 3	
592	1fgv	L	25	114	1.4e-45	0.21	0.53		IMMUNOGLOBULIN FV	

Table 5

SEQ ID NO:	PDB ID.	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 IFGV 3 ANTIBODY 'H52' (HUH52-AA FV) 1FGV 4	
592	2fb4	L	26	117	1.2e-44	0.35	0.82		IMMUNOGLOBULIN IMMUNOGLOBULIN FAB 2FB4 4	
592	2fgw	L	25	114	1.7e-45	0.32	0.77		IMMUNOGLOBULIN FAB FRAGMENT OF A HUMANIZED VERSION OF THE ANTI-CD18 2FGW 3 ANTIBODY 'H52' (HUH52-OZ FAB) 2FGW 4	
603	1cru	A	169	400	1.5e-46	0.13	0.28		SOLUBLE QUINOPROTEIN GLUCOSE DEHYDROGENASE; CHAIN: A, B;	OXIDOREDUCTASE BETA-PROPELLER, SUPERBARREL, COMPLEX WITH THE COFACTOR PQQ 2 AND THE INHIBITOR METHYLHYDRAZINE, OXIDOREDUCTASE
603	1cru	A	186	404	7.5e-49	0.01	0.27		SOLUBLE QUINOPROTEIN GLUCOSE DEHYDROGENASE; CHAIN: A, B;	OXIDOREDUCTASE BETA-PROPELLER, SUPERBARREL, COMPLEX WITH THE COFACTOR PQQ 2 AND THE INHIBITOR METHYLHYDRAZINE, OXIDOREDUCTASE
615	1c0t	A	174	431	1.7e-65	-0.24	0.25		HIV-1 REVERSE	TRANSFERASE HIV-1 REVERSE

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
615	1c0t	B	176	431	1e-62	-0.31	0.23		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
615	1c1c	B	175	431	1e-74	-0.12	0.39		HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN); CHAIN: B;	TRANSFERASE HIV-1 REVERSE TRANSCRIPTASE, AIDS, NON-NUCLEOSIDE INHIBITOR, 2 DRUG DESIGN
615	1c9r	A	171	431	1e-70	-0.08	0.94		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'- CHAIN: T; DNA (5'- CHAIN: P;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184ILE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA
615	1c9r	B	171	431	1.7e-79	-0.14	0.59		HIV-1 REVERSE TRANSCRIPTASE (CHAIN A); CHAIN: A; HIV-1 REVERSE TRANSCRIPTASE (CHAIN B); CHAIN: B; ANTIBODY (LIGHT CHAIN); CHAIN: L; ANTIBODY (HEAVY CHAIN); CHAIN: H; DNA (5'- CHAIN: T; DNA (5'- CHAIN: P;	TRANSFERASE/IMMUNE SYSTEM/DNA HIV-1 RT; HIV-1 RT; HIV, REVERSE TRANSCRIPTASE, MET184ILE, 3TC, PROTEIN-DNA 2 COMPLEX, DRUG RESISTANCE, M184I, TRANSFERASE/IMMUNE 3 SYSTEM/DNA

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
615	1mml		154	396	5.1e-50			116.10	MMLV REVERSE TRANSCRIPTASE; 1MML 4	REVERSE TRANSCRIPTASE
615	1rth	A	171	431	3.4e-86	-0.12	0.74		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
615	1rth	B	173	431	1e-75	-0.09	0.23		HIV-1 REVERSE TRANSCRIPTASE; 1RTH 4 CHAIN: A, B; 1RTH 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1RTH 6 HIV-1 REVERSE TRANSCRIPTASE 1RTH 15
615	1vrt	A	174	431	1.7e-85	-0.26	0.40		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
615	1vrt	B	175	431	1.7e-75	-0.20	0.11		HIV-1 REVERSE TRANSCRIPTASE; 1VRT 4 CHAIN: A, B; 1VRT 5	NUCLEOTIDYLTRANSFERASE HIV-1 RT; 1VRT 6 HIV-1 REVERSE TRANSCRIPTASE 1VRT 15
615	3hvt	B	172	431	3.4e-74	-0.14	0.00		NUCLEOTIDYLTRANSFERASE REVERSE TRANSCRIPTASE (E.C.2.7.7.49) 3HVT 3	
620	1aut	L	47	75	0.00068	-0.18	0.42		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										COAGULATION/INHIBITOR)
620	1diy	A	46	77	0.00068	0.69	0.25		PROSTAGLANDIN H2 SYNTHASE-1; CHAIN: A;	OXIDOREDUCTASE ARACHIDONIC ACID, MEMBRANE PROTEIN, PEROXIDASE, DIOXYGENASE
620	1fsb		46	75	0.0034	1.08	0.34		P-SELECTIN; CHAIN: NULL;	CELL ADHESION PROTEIN EGF-LIKE DOMAIN, CELL ADHESION PROTEIN, TRANSMEMBRANE, 2 GLYCOPROTEIN
627	1mgl	A	260	376	3.4e-28	-0.94	0.06		HTLV-1 GP21 ECTODOMAIN/MALTOSE-BINDING PROTEIN CHAIN: A;	LEUKEMIA VIRUS TYPE 1 HUMAN T CELL LEUKEMIA VIRUS TYPE 1, HTLV-1, ENVELOPE 2 PROTEIN, MEMBRANE FUSION, MALTOSE-BINDING PROTEIN CHIMERA
627	2ebo	A	304	376	5.1e-22	-0.56	0.21		EBOLA VIRUS ENVELOPE GLYCOPROTEIN; CHAIN: A, B, C;	ENVELOPE GLYCOPROTEIN ENVELOPE GLYCOPROTEIN, FILOVIRUS, EBOLA VIRUS, GP2, COAT 2 PROTEIN
658	1a9n	A	27	164	2.5e-18	0.22	0.69		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B''; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP,RIBONUCLEOPROTEIN
658	1a9n	A	54	188	5e-24	0.30	0.48		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B''; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP,RIBONUCLEOPROTEIN
658	1a9n	C	27	164	7.5e-18	0.38	0.96		U2 RNA HAIRPIN IV; CHAIN:	COMPLEX (NUCLEAR

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP,RIBONUCLEOPROTEIN
658	1a9n	C	54	188	1.5e-23	0.46	0.53		U2 RNA HAIRPIN IV; CHAIN: Q, R; U2 A'; CHAIN: A, C; U2 B"; CHAIN: B, D;	COMPLEX (NUCLEAR PROTEIN/RNA) COMPLEX (NUCLEAR PROTEIN/RNA), RNA, SNRNP,RIBONUCLEOPROTEIN
658	1d0b	A	70	237	1.7e-21	-0.00	0.41		INTERNALIN B; CHAIN: A;	CELL ADHESION LEUCINE RICH REPEAT, CALCIUM BINDING, CELL ADHESION
658	1dce	A	98	218	1.2e-09	-0.43	0.30		RAB GERANYLGERANYLTRANSFERASE ALPHA SUBUNIT; CHAIN: A, C; RAB GERANYLGERANYLTRANSFERASE BETA SUBUNIT; CHAIN: B, D;	TRANSFERASE CRYSTAL STRUCTURE, RAB GERANYLGERANYLTRANSFERASE, 2.0 A 2 RESOLUTION, N-FORMYLMETHIONINE, ALPHA SUBUNIT, BETA SUBUNIT
658	1ds9	A	55	178	2.5e-17	-0.29	0.06		OUTER ARM DYNEIN; CHAIN: A;	CONTRACTILE PROTEIN LEUCINE-RICH REPEAT, BETA-BETA-ALPHA CYLINDER, DYNEIN, 2 CHLAMYDOMONAS, FLAGELLA
658	2bnh		34	183	1e-21	0.28	-0.03		RIBONUCLEASE INHIBITOR; CHAIN: NULL;	ACETYLATION RNASE INHIBITOR, RIBONUCLEASE/ANGIOGENIN INHIBITOR ACETYLATION, LEUCINE-RICH REPEATS
665	1a0h	A	30	150	2.5e-29	0.37	0.65		MEIZOTHROMBIN; CHAIN: A, B, D, E; D-PHE-PRO-ARG; CHAIN: C, F;	COMPLEX (SERINE PROTEASE/INHIBITOR) DESF1; PPACK; SERINE PROTEASE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										COAGULATION, THROMBIN, PROTHROMBIN, 2 MEIZOTHROMBIN, COMPLEX (SERINE PROTEASE/INHIBITOR)
665	1a0h	A	30	169	6.8e-10	0.28	0.76		MEIZOTHROMBIN; CHAIN: A, B, D, E; D-PHE-PRO-ARG; CHAIN: C, F;	COMPLEX (SERINE PROTEASE/INHIBITOR) DESF1; PPACK; SERINE PROTEASE, COAGULATION, THROMBIN, PROTHROMBIN, 2 MEIZOTHROMBIN, COMPLEX (SERINE PROTEASE/INHIBITOR)
665	1a0h	A	30	201	2.5e-29			82.71	MEIZOTHROMBIN; CHAIN: A, B, D, E; D-PHE-PRO-ARG; CHAIN: C, F;	COMPLEX (SERINE PROTEASE/INHIBITOR) DESF1; PPACK; SERINE PROTEASE, COAGULATION, THROMBIN, PROTHROMBIN, 2 MEIZOTHROMBIN, COMPLEX (SERINE PROTEASE/INHIBITOR)
665	1b2i	A	32	120	7.5e-26			72.58	PLASMINOGEN; CHAIN: A;	HYDROLASE SERINE PROTEASE, FIBRINOLYSIS, LYSINE-BINDING DOMAIN, 2 PLASMINOGEN, KRINGLE 2, HYDROLASE
665	1b2i	A	34	119	7.5e-26	0.90	0.81		PLASMINOGEN; CHAIN: A;	HYDROLASE SERINE PROTEASE, FIBRINOLYSIS, LYSINE-BINDING DOMAIN, 2 PLASMINOGEN, KRINGLE 2, HYDROLASE
665	1cea	A	35	119	1e-24			68.58	PLASMINOGEN; 1CEA 7 CHAIN: A, B; 1CEA 8	SERINE PROTEASE KIPG; 1CEA 10
665	1kdu		35	120	2.5e-28			71.09	PLASMINOGEN ACTIVATION	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									PLASMINOGEN ACTIVATOR (UROKINASE-TYPE, KRINGLE DOMAIN) 1KDU 3 (U-PA K) (NMR, MINIMIZED AVERAGE STRUCTURE) 1KDU 4	
665	1kdu		36	119	2.5e-28	0.91	0.96		PLASMINOGEN ACTIVATION PLASMINOGEN ACTIVATOR (UROKINASE-TYPE, KRINGLE DOMAIN) 1KDU 3 (U-PA K) (NMR, MINIMIZED AVERAGE STRUCTURE) 1KDU 4	
665	1krn		35	119	5e-22			76.76	PLASMINOGEN; CHAIN: NULL;	SERINE PROTEASE KRINGLE, BLOOD, PLASMINOGEN, SERINE PROTEASE
665	1pml	A	34	119	1.3e-28	0.89	1.00		HYDROLASE(SERINE PROTEASE) TISSUE PLASMINOGEN ACTIVATOR KRINGLE 2 (E.C.3.4.21.68) 1PML 3	
665	1pml	A	34	121	1.3e-28			86.47	HYDROLASE(SERINE PROTEASE) TISSUE PLASMINOGEN ACTIVATOR KRINGLE 2 (E.C.3.4.21.68) 1PML 3	
665	1pml	C	34	119	1e-28	0.94	0.96		HYDROLASE(SERINE PROTEASE) TISSUE PLASMINOGEN ACTIVATOR KRINGLE 2 (E.C.3.4.21.68) 1PML 3	
665	1pml	C	34	120	1e-28			86.67	HYDROLASE(SERINE PROTEASE) TISSUE	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									PLASMINOGEN ACTIVATOR KRINGLE 2 (E.C.3.4.21.68) 1PML 3	
665	1sfp		218	329	2.5e-17	1.13	0.99		ASFP; CHAIN: NULL;	SPERMADHESIN ACIDIC SEMINAL PROTEIN; SPERMADHESIN, BOVINE SEMINAL PLASMA PROTEIN, ACIDIC 2 SEMINAL FLUID PROTEIN, ASFP, CUB DOMAIN, X-RAY CRYSTAL 3 STRUCTURE, GROWTH FACTOR
665	1sfp		238	327	3.4e-07	0.62	0.09		ASFP; CHAIN: NULL;	SPERMADHESIN ACIDIC SEMINAL PROTEIN; SPERMADHESIN, BOVINE SEMINAL PLASMA PROTEIN, ACIDIC 2 SEMINAL FLUID PROTEIN, ASFP, CUB DOMAIN, X-RAY CRYSTAL 3 STRUCTURE, GROWTH FACTOR
665	1spp	A	218	323	2.5e-16	0.67	0.11		MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-I; CHAIN: A; MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II; CHAIN: B	COMPLEX (SEMINAL PLASMA PROTEIN/SPP) SEMINAL PLASMA PROTEINS, SPERMADHESINS, CUB DOMAIN 2 ARCHITECTURE, COMPLEX (SEMINAL PLASMA PROTEIN/SPP)
665	1spp	B	218	323	2.5e-15	0.62	-0.07		MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-I; CHAIN: A; MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II; CHAIN: B	COMPLEX (SEMINAL PLASMA PROTEIN/SPP) SEMINAL PLASMA PROTEINS, SPERMADHESINS, CUB DOMAIN 2 ARCHITECTURE, COMPLEX (SEMINAL PLASMA

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										PROTEIN/SPP)
665	1spp	B	245	328	6.8e-06	0.38	0.09		MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-I; CHAIN: A; MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II; CHAIN: B	COMPLEX (SEMINAL PLASMA PROTEIN/SPP) SEMINAL PLASMA PROTEINS, SPERMADHESINS, CUB DOMAIN 2 ARCHITECTURE, COMPLEX (SEMINAL PLASMA PROTEIN/SPP)
665	1urk		1	123	2.2e-23			69.71	PLASMINOGEN ACTIVATION PLASMINOGEN ACTIVATOR (UROKINASE-TYPE) (AMINO TERMINAL FRAGMENT) (NMR, 15 STRUCTURES)	
665	2hpp	P	36	119	5e-25			66.47	HYDROLASE(SERINE PROTEINASE) ALPHA-THROMBIN (E.C.3.4.21.5) COMPLEX WITH 2HPP 3 D-PHE-PRO-ARG-CHLOROMETHYLKETONE (PPACK) CHLOROMETHYLKETONE 2HPP 4 REPLACED BY A METHYLENE GROUP AND BOVINE PROTHROMBIN 2HPP 5 FRAGMENT 2 2HPP 6	
665	2hpp	P	36	119	5e-25	0.71	0.39		HYDROLASE(SERINE PROTEINASE) ALPHA-THROMBIN (E.C.3.4.21.5) COMPLEX WITH 2HPP 3 D-PHE-PRO-ARG-CHLOROMETHYLKETONE (PPACK)	

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
									CHLOROMETHYLKETONE 2HPP 4 REPLACED BY A METHYLENE GROUP AND BOVINE PROTHROMBIN 2HPP 5 FRAGMENT 2 2HPP 6	
665	2hpq	P	36	119	1.2e-24			60.25	HYDROLASE(SERINE PROTEINASE) ALPHA-THROMBIN (E.C.3.4.21.5) COMPLEX WITH 2HPQ 3 D-PHE-PRO-ARG-CHLOROMETHYLKETONE (PPACK) CHLOROMETHYLKETONE 2HPQ 4 REPLACED BY A METHYLENE GROUP AND HUMAN PROTHROMBIN 2HPQ 5 FRAGMENT 2 2HPQ 6	
665	2pf1		20	119	2.3e-25	0.85	0.71		HYDROLASE(SERINE PROTEINASE) PROTHROMBIN FRAGMENT 1 (RESIDUES 1 - 156) 2PF1 3	
665	2pf1		5	131	2.3e-25			58.78	HYDROLASE(SERINE PROTEINASE) PROTHROMBIN FRAGMENT 1 (RESIDUES 1 - 156) 2PF1 3	
665	2pf2		35	119	2.5e-25	0.82	0.77		HYDROLASE(SERINE PROTEASE) PROTHROMBIN FRAGMENT 1 (RESIDUES 1 - 156) COMPLEX WITH 2PF2 3 CALCIUM 2PF2 4	
665	3kiv		35	119	5e-27			76.56	APOLIPOPROTEIN; CHAIN: NULL;	KRINGLE KRINGLE, LYSINE BINDING SITE,

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Compound	PDB annotation
										APOLIPOPROTEIN(A)
665	3kiv		35	119	5e-27	0.76	0.87		APOLIPOPROTEIN; CHAIN: NULL;	KRINGLE KRINGLE, LYSINE BINDING SITE, APOLIPOPROTEIN(A)
665	5hpg	A	35	122	1e-26			77.19	PLASMINOGEN; CHAIN: A, B;	SERINE PROTEASE SERINE PROTEASE, KRINGLE 5, HUMAN PLASMINOGEN, FIBRINOLYSIS
665	5hpg	A	35	122	1e-26	0.65	0.70		PLASMINOGEN; CHAIN: A, B;	SERINE PROTEASE SERINE PROTEASE, KRINGLE 5, HUMAN PLASMINOGEN, FIBRINOLYSIS
665	9wga	A	21	168	1.7e-13	0.15	-0.12		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
665	9wga	A	51	234	3.4e-10	0.16	-0.19		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	

Table 6
264

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
337	29	0.968	0.793
338	32	0.989	0.841
339	37	0.972	0.775
341	42	0.943	0.626
342	34	0.993	0.933
344	33	0.968	0.827
345	28	0.995	0.945
346	26	0.994	0.932
347	41	0.959	0.629
348	39	0.986	0.641
349	28	0.988	0.935
350	24	0.981	0.776
351	25	0.898	0.612
352	14	0.943	0.864
353	24	0.976	0.925
355	21	0.896	0.706
357	39	0.983	0.710
358	26	0.971	0.899
359	27	0.970	0.898
360	27	0.970	0.898
362	29	0.964	0.562
363	33	0.937	0.698
364	24	0.988	0.952
365	18	0.995	0.978
366	13	0.972	0.733
367	25	0.992	0.929
368	20	0.987	0.963
369	41	0.972	0.714
370	40	0.993	0.805
372	40	0.993	0.805
373	42	0.890	0.551
375	21	0.942	0.816
376	25	0.954	0.816
378	41	0.983	0.859
379	25	0.980	0.906
380	15	0.953	0.860
381	22	0.943	0.746
382	31	0.995	0.895
383	17	0.959	0.867
385	18	0.981	0.858
387	22	0.993	0.966
388	49	0.987	0.594
390	25	0.990	0.857
391	26	0.985	0.956
392	19	0.993	0.953
393	48	0.985	0.571
394	17	0.976	0.772
395	15	0.932	0.796
396	40	0.996	0.972
398	25	0.941	0.656
399	16	0.984	0.949
401	34	0.971	0.910
402	42	0.983	0.683
403	17	0.961	0.884
405	17	0.961	0.884

Table 6
265

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
406	26	0.996	0.922
407	20	0.947	0.881
408	48	0.940	0.755
409	30	0.968	0.777
410	32	0.953	0.778
411	20	0.963	0.551
412	25	0.958	0.928
414	33	0.988	0.893
415	24	0.933	0.671
416	44	0.956	0.803
417	47	0.967	0.826
418	48	0.992	0.807
419	25	0.976	0.909
421	29	0.973	0.792
422	29	0.922	0.662
423	32	0.967	0.646
424	21	0.933	0.785
425	31	0.894	0.613
426	46	0.981	0.714
427	44	0.955	0.611
428	17	0.950	0.712
429	14	0.989	0.917
430	27	0.998	0.952
431	35	0.969	0.716
432	17	0.943	0.681
433	21	0.956	0.879
434	25	0.985	0.718
435	17	0.943	0.794
436	29	0.998	0.924
437	29	0.998	0.924
438	21	0.986	0.966
442	25	0.988	0.947
443	18	0.900	0.591
444	23	0.975	0.884
445	18	0.898	0.719
446	43	0.907	0.701
447	29	0.941	0.708
448	20	0.989	0.960
449	20	0.989	0.960
450	40	0.998	0.990
451	35	0.984	0.757
452	42	0.977	0.671
453	15	0.978	0.902
454	17	0.976	0.927
455	34	0.957	0.706
456	18	0.978	0.937
459	18	0.902	0.649
460	36	0.978	0.657
461	19	0.973	0.788
462	20	0.964	0.774
463	24	0.978	0.709
464	21	0.968	0.782
465	45	0.998	0.924
466	22	0.989	0.960
467	49	0.986	0.825

Table 6
266

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
468	38	0.959	0.769
469	28	0.988	0.744
470	24	0.909	0.643
471	20	0.972	0.830
472	48	0.957	0.617
473	20	0.980	0.902
474	17	0.905	0.697
475	47	0.995	0.684
477	20	0.983	0.888
478	31	0.977	0.806
481	38	0.930	0.725
482	20	0.972	0.888
483	10	0.993	0.569
484	34	0.994	0.867
485	23	0.904	0.643
486	22	0.974	0.877
487	17	0.959	0.814
488	48	0.946	0.768
490	19	0.957	0.838
491	38	0.988	0.950
492	24	0.967	0.918
494	31	0.945	0.695
495	46	0.992	0.562
496	23	0.958	0.866
497	25	0.973	0.888
498	41	0.981	0.577
499	43	0.970	0.727
500	32	0.913	0.607
501	27	0.962	0.882
502	22	0.989	0.887
503	22	0.981	0.881
504	28	0.972	0.825
505	31	0.990	0.766
506	30	0.995	0.964
507	24	0.955	0.640
508	37	0.977	0.860
509	38	0.983	0.775
510	18	0.990	0.922
511	24	0.993	0.923
512	22	0.948	0.754
513	22	0.989	0.927
514	41	0.987	0.895
515	31	0.979	0.864
516	16	0.988	0.968
518	27	0.977	0.934
519	43	0.994	0.918
520	45	0.995	0.686
522	26	0.975	0.807
523	30	0.982	0.647
524	42	0.982	0.664
525	15	0.935	0.811
526	36	0.999	0.992
528	41	0.901	0.614
529	20	0.994	0.976
530	21	0.940	0.738

Table 6
267

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
531	38	0.991	0.889
532	16	0.915	0.719
534	28	0.974	0.886
535	21	0.981	0.911
536	30	0.993	0.832
537	30	0.993	0.832
538	21	0.993	0.979
539	38	0.884	0.655
540	25	0.963	0.849
541	27	0.954	0.863
542	27	0.961	0.767
544	19	0.972	0.877
546	25	0.986	0.802
547	45	0.954	0.577
548	26	0.895	0.712
549	23	0.956	0.836
550	19	0.989	0.950
551	40	0.967	0.821
552	19	0.968	0.923
553	44	0.990	0.566
554	41	0.922	0.748
555	34	0.991	0.758
557	32	0.968	0.678
558	23	0.989	0.965
559	23	0.989	0.965
560	16	0.969	0.917
561	19	0.978	0.930
562	39	0.982	0.678
563	36	0.987	0.866
564	24	0.942	0.780
565	46	0.963	0.617
567	49	0.998	0.716
568	45	0.996	0.966
569	32	0.971	0.914
570	25	0.998	0.958
571	25	0.998	0.958
573	41	0.962	0.555
574	19	0.973	0.893
575	37	0.968	0.621
576	24	0.983	0.949
577	40	0.980	0.824
578	21	0.953	0.854
580	45	0.987	0.852
581	18	0.898	0.665
583	24	0.959	0.869
584	20	0.982	0.852
585	44	0.894	0.594
586	48	0.981	0.692
588	17	0.992	0.969
590	29	0.975	0.835
591	17	0.924	0.748
592	25	0.974	0.872
593	18	0.943	0.843
594	33	0.970	0.887
595	25	0.980	0.893

Table 6
268

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
596	18	0.973	0.922
597	26	0.994	0.969
598	34	0.961	0.562
599	39	0.978	0.791
600	17	0.928	0.753
603	19	0.976	0.950
605	49	0.994	0.792
606	24	0.993	0.937
607	19	0.991	0.956
608	39	0.996	0.930
611	43	0.987	0.765
612	41	0.977	0.722
613	23	0.952	0.651
615	19	0.987	0.898
617	26	0.972	0.732
618	20	0.965	0.833
619	13	0.923	0.755
620	25	0.951	0.738
622	30	0.967	0.769
623	48	0.979	0.568
625	18	0.956	0.655
626	27	0.975	0.831
627	44	0.987	0.725
628	35	0.969	0.616
629	33	0.981	0.884
630	35	0.954	0.759
631	20	0.926	0.787
632	20	0.974	0.908
633	16	0.888	0.686
635	27	0.973	0.870
636	37	0.956	0.698
637	25	0.969	0.873
638	48	0.985	0.705
640	26	0.956	0.717
641	11	0.977	0.958
642	22	0.953	0.916
643	39	0.972	0.817
644	29	0.983	0.897
645	24	0.917	0.657
646	23	0.967	0.856
648	25	0.928	0.667
650	38	0.966	0.856
651	21	0.990	0.950
652	41	0.971	0.804
653	19	0.937	0.870
654	15	0.987	0.802
655	20	0.925	0.699
657	40	0.977	0.661
658	14	0.967	0.876
659	41	0.990	0.724
660	23	0.968	0.924
661	27	0.882	0.585
662	44	0.990	0.644
664	17	0.950	0.658
665	25	0.971	0.897

Table 6
269

SEQ ID NO:	Position of Signal Peptide	Maximum score	Average score
666	39	0.996	0.868
667	20	0.987	0.946
669	14	0.946	0.864
672	26	0.982	0.896

Table 7
270

SEQ ID NO:	Chromsomal location
1	5
2	6
3	3
4	17
5	11
6	10
7	3q
8	13
9	17
10	1
12	13
13	17
15	13
16	5
17	6
19	10
21	1
23	13
24	1
25	11
26	12
28	12
29	15
30	19
31	1
32	11
33	1p31.2-32.3
35	7
36	17
37	8
39	10
40	7
41	3
42	22
43	12
44	13
45	13
46	17
47	20
48	16
49	11
50	15
51	15
52	15
53	15
54	15
56	16
57	15
58	22.
60	1
61	1
62	7
63	11
66	2
67	8

Table 7

271

SEQ ID NO:	Chromsomal location
68	16
69	1
70	11
71	13
72	18
75	12
76	10
77	9
78	17
80	3
81	11
82	8
84	2
85	11
86	4
87	9
88	9
89	8
91	16
92	15
93	4
94	19
95	7
96	16
97	15
98	16
99	5
100	19
101	8
102	1p35.1-35.3.
104	6
105	8
106	5
107	1p34.1-36.11
109	4
110	6
111	1
112	19
113	6
114	10
115	18
116	6
117	19
118	1
119	3
120	3
121	13
122	3
123	12
124	1
125	6
126	13
128	5
129	16
130	4

Table 7

272

SEQ ID NO:	Chromsomal location
131	5
133	5
134	12
137	Xq25-26
138	13
140	6p11.2-12.3
141	19
142	6q16-21
143	1q23-24.
144	5
145	22.
146	17
147	16
148	5
151	19
152	17
153	16
154	18
155	5
156	10
157	2
158	9
159	9
160	20
161	17
162	17
163	5
165	20
166	6
167	16
168	12
170	6
172	9
173	3
174	20
175	16
176	11
177	18
178	10
179	22q13.1-13.2.
181	16
182	11
183	17
184	7
185	11
186	06
187	6q16.2-21
188	3
190	19
191	19
192	1
193	16
194	X
195	6
197	Xp11.4-21.2

Table 7

273

SEQ ID NO:	Chromsomal location
198	1
199	8
200	6q22.1-22.33
201	8
204	6
206	17
207	19
210	17
211	8
212	15
213	15
214	11
215	15
216	15
218	1
219	2
220	12
222	6q23.1-24.3
224	16
225	21
226	15
227	1
228	17
229	1
232	16
234	3
235	22
236	10
237	3
238	16
239	3
240	17
241	2
242	3
243	13
244	13
246	17
247	15
250	1
251	5
252	19
253	9
255	3
256	14
257	15
258	1
259	3
260	16
262	X
263	11
264	21
265	3
266	3
267	14
270	11

Table 7
274

SEQ ID NO:	Chromsomal location
272	17
273	15
276	18
277	4
278	17
279	1
280	6.
281	22q13.1-13.33
282	3
284	20
286	6
288	4
290	1
291	1
292	1
293	1
294	11
295	9
296	3
298	1
300	1
301	11
302	6p21.1-21.2
303	17
304	3
305	12
306	16
307	5
309	17
312	5
313	18
314	16
315	18
316	11
317	5
318	1q42.2-43
319	11
320	19
321	3
324	3
326	5
327	8
329	16
330	4
332	6
333	12
334	12
335	18
336	2

Table 8
275

SEQ ID NO:	Number of Transmembrane Domains Predicted	Position of Transmembrane Region: TMPred Score
337	1	9-31:2958
338	1	15-38:1948
339	2	20-34:1518 82-98:1908
340	1	64-80:1560
341	1	24-40:2347
342	1	14-32:2720
343	1	23-44:1807
344	2	15-31:1300 118-140:3012
345	4	95-111:2524 104-139:1338 125-147:2138 174-209:1036
346	2	6-38:1711 49-67:1103
347	2	15-31:3431 69-86:889
348	1	28-44:2183
349	3	13-32:2547 95-110:1692 112-132:1903
353	3	41-57:1768 82-97:2647 122-136:968
354	1	250-265:1867
355	3	46-62:911 68-84:1367 154-166:1297
356	2	32-51:2342 114-130:1188
357	1	23-39:2309
359	1	41-59:2412
360	2	85-114:2984 221-238:959
361	2	35-50:1595 66-85:2779
362	2	17-32:1331 57-71:1728
363	3	14-31:1963 40-58:1009 66-86:1248
364	1	226-242:2202
365	2	46-61:832 73-90:2191
366	1	34-56:1058
367	1	154-172:2074
368	3	34-49:1210 66-99:1252 97-113:2355
369	1	18-33:1975
370	2	34-53:1125 67-84:2061
371	4	158-174:1945 199-216:1112 225-242:1673 254-271:946
372	1	15-33:1775
373	1	181-199:1868
374	5	38-54:1712 67-94:2110 114-128:918 240-256:855 277-292:1359
375	2	50-74:2625 130-149:1166
376	4	16-38:1473 43-59:1371 77-94:1851 199-214:1092
377	1	46-62:3051
378	1	17-34:2743
379	1	95-118:3033
380	1	213-230:985
382	1	8-31:3667
383	1	83-101:2361
384	3	47-62:1204 51-79:1625 96-109:1118
386	4	13-35:1282 58-73:2648 91-107:1319 148-165:1783
387	4	41-56:1354 62-78:1639 88-103:977 134-150:1946
388	2	25-46:2369 66-81:1705
389	5	20-43:823 51-73:1163 87-106:1827 105-125:1017 153-186:1554
391	1	74-89:3414
393	1	31-57:2521
394	3	27-46:2157 130-160:1822 236-250:888

Table 8

276

SEQ ID NO:	Number of Transmembrane Domains Predicted	Position of Transmembrane Region: TMPred Score
396	10	28-44:2267 50-76:1625 68-88:2769 93-113:1629 118-138:2697 153-168:1629 178-194:2313 203-238:1733 244-263:2730 269-284:1367
397	1	40-67:1986
400	3	23-40:2163 266-285:985 291-304:1229
401	3	18-34:2249 256-272:1362 280-299:1671
402	1	21-39:2045
403	2	34-51:1665 133-151:1190
404	4	21-37:2440 57-74:1286 84-112:1585 122-143:1004
405	2	48-63:1829 197-216:1112
408	1	29-48:1619
410	2	16-32:1602 191-205:890
411	3	44-60:2409 103-123:941 165-185:2002
413	3	19-35:2153 38-53:1100 78-97:1064
414	1	20-39:1830
415	2	57-72:2060 93-110:939
416	8	23-47:1290 60-80:1779 87-106:1447 159-187:2236 202-216:1085 234-249:981 270-299:1491 324-338:1352
417	1	21-39:2481
418	2	27-52:1562 66-84:864
419	2	15-31:1529 41-56:2722
420	1	21-36:2544
421	2	16-34:1960 40-55:951
422	1	174-191:1728
423	1	16-32:827
424	1	45-66:1964
425	1	75-92:1800
426	3	17-40:2165 71-83:1112 116-143:1198
427	1	23-39:3165
428	1	42-59:859
430	5	75-90:1359 107-122:1520 135-151:1967 175-191:1416 236-251:2332
431	1	14-32:2317
435	2	214-236:1046 282-294:966
436	1	48-63:2723
437	6	125-141:2144 157-173:1116 185-204:1756 223-238:926 243-259:1271 273-288:1225
438	2	38-55:1680 151-168:2550
439	2	30-51:2155 161-176:905
440	6	36-50:2210 58-74:1644 126-141:914 152-173:1406 187-202:2224 221-236:1055
441	5	49-70:1075 88-104:1052 123-140:1710 157-175:2590 191-204:1390
442	2	25-45:1365 64-84:1812
444	2	46-59:1059 186-206:1046
445	1	97-112:1026
446	1	26-41:1887
448	3	28-43:1680 58-73:1675 90-105:1928
449	3	82-102:1765 119-134:1405 167-183:2521
450	4	15-45:1726 42-67:2522 207-229:861 274-291:922
451	1	13-31:2843
452	3	23-38:1889 50-66:831 121-137:1096
453	3	19-35:1356 72-87:1830 105-120:1373

Table 8
277

SEQ ID NO:	Number of Transmembrane Domains Predicted	Position of Transmembrane Region: TMPred Score
455	2	22-48:1148 384-399:2339
457	1	36-51:2076
458	6	83-100:2781 111-133:1847 157-173:2151 175-191:1172 236-251:3053 307-322:1307
460	1	14-34:2733
461	1	31-50:2047
462	1	118-137:812
464	1	234-248:948
465	1	7-41:2396
467	1	18-33:1771
468	1	15-39:2946
470	1	53-68:3633
471	1	36-51:1750
472	3	30-58:2255 69-85:1303 102-116:965
473	7	5-33:2407 48-62:834 82-101:1768 116-136:1635 165-185:2884 226-247:1338 263-282:1779
475	1	26-47:2958
476	1	43-58:2185
477	1	51-66:896
478	1	20-39:1851
479	1	30-48:2719
480	2	50-67:1746 105-120:1144
481	1	142-159:2140
482	1	108-123:1623
483	1	34-48:2268
484	2	14-38:2868 281-297:941
486	1	217-239:1272
487	1	146-168:2684
488	2	90-107:1944 363-377:1338
489	3	64-81:2157 84-100:1243 97-133:1672
490	1	48-72:2661
491	3	2-38:971 22-46:1497 84-99:1261
493	2	34-61:2058 93-108:1716
494	2	40-59:1918 234-249:859
495	1	24-45:2330
497	1	296-313:812
498	1	21-44:2763
499	1	21-36:2617
500	1	26-51:825
502	4	34-55:2354 150-169:1592 311-333:1867 353-375:892
503	1	69-87:2593
504	5	59-80:1228 88-107:866 157-176:3161 198-216:1250 223-238:2194
505	1	195-210:1193
506	1	19-35:2865
507	1	69-98:822
508	3	18-33:2344 94-115:1093 232-249:1415
509	1	14-31:2117
510	1	166-182:2113
514	1	17-35:2291
515	1	11-31:871
517	1	31-53:2985

Table 8
278

SEQ ID NO:	Number of Transmembrane Domains Predicted	Position of Transmembrane Region: TMPred Score
519	1	20-44:2459
520	1	20-37:2284
521	1	22-42:3116
522	1	46-62:2496
524	1	19-33:1834
526	2	41-71:1782 65-86:3101
527	3	19-34:1101 46-62:1928 185-201:1841
528	1	17-39:1978
529	1	364-379:1065
531	1	22-40:1765
533	1	38-53:1788
534	1	14-32:2099
535	2	32-52:1769 77-102:2317
536	3	16-37:895 52-69:1796 100-120:1617
537	4	153-175:2138 189-204:1068 261-283:2271 290-306:1112
538	1	1-34:1975
539	1	10-38:1023
540	1	15-31:1522
541	1	74-91:2543
542	5	49-64:1187 82-96:1485 119-140:1408 129-153:2110 206-222:2257
543	1	66-83:2200
546	2	75-94:924 180-195:1494
547	1	22-37:2183
548	5	43-67:2282 70-91:1282 121-137:2440 169-183:1439 197-232:1120
549	3	14-34:1791 83-97:1381 115-144:1592
550	4	43-62:1533 195-216:2160 222-237:1314 257-270:1867
551	2	13-31:1516 69-88:2277
552	5	25-42:1555 74-89:1237 114-142:2195 154-169:1023 185-200:2114
553	3	24-47:1711 61-79:2020 192-207:2454
554	2	36-56:1076 90-110:1216
555	1	16-33:2206
556	2	17-36:2654 64-76:932
557	1	19-34:1366
558	2	21-46:1142 54-70:3147
560	1	28-46:2247
561	2	23-43:1069 58-75:1756
562	4	21-39:1494 81-97:1518 125-143:1312 148-169:2440
563	10	7-32:2014 82-96:1124 107-123:1475 148-167:1298 170-193:1565 258-273:1090 296-316:1839 324-345:1356 354-369:1159 420-437:1669
564	2	44-60:963 75-90:3007
565	4	29-44:1865 76-93:1315 119-138:1894 155-176:1330
566	1	42-69:2215
567	2	36-55:2620 41-76:845
568	1	3-35:3176
569	1	56-73:3062
572	3	45-61:2010 110-125:1024 175-193:839
573	1	18-39:2254
574	3	55-76:2276 89-112:1167 148-168:2134

Table 8
279

SEQ ID NO:	Number of Transmembrane Domains Predicted	Position of Transmembrane Region: TMPred Score
575	1	16-36:2701
576	2	82-107:1813 168-186:2844
577	1	17-35:2449
578	1	36-53:2305
579	1	29-45:2349
580	1	26-43:2340
581	2	238-257:908 396-412:1281
582	2	50-68:1787 82-94:808
583	2	41-55:1214 76-91:2379
584	1	120-139:1924
585	2	25-41:2077 208-223:986
586	2	25-45:1955 167-181:1187
587	3	47-62:2783 76-92:1090 115-130:2791
589	1	58-85:1106
590	4	33-48:1166 71-88:2044 108-123:1229 134-154:2709
593	1	79-94:1909
594	6	16-33:2461 94-113:2485 137-152:1212 190-212:3236 237-253:971 266-285:1138
596	2	48-66:1420 56-86:2350
597	1	14-32:2650
598	2	23-42:2154 134-155:1123
599	3	16-34:1811 55-70:1301 82-99:1627
600	1	43-58:890
601	1	27-42:2043
602	3	52-75:2018 325-346:865 375-392:839
603	1	353-370:2096
604	1	25-45:2047
605	1	24-47:2800
606	2	71-86:1595 102-121:2779
607	1	297-319:2854
608	10	25-41:1489 54-72:2563 87-103:1436 116-134:2525 149-165:1474 178-196:2516 211-227:1420 240-258:2456 273-289:1392 302-320:2395
609	2	22-48:2007 141-164:1410
610	2	21-41:1941 102-117:3056
611	8	29-44:1389 61-74:917 88-103:1267 115-129:890 179-193:898 204-221:1978 220-238:1076 259-275:1735
612	1	26-43:1767
614	2	36-51:2233 100-113:2408
615	2	40-56:1175 69-85:1803
619	1	35-53:2023
621	4	17-32:2238 39-60:1679 79-95:2605 114-129:1098
623	1	23-42:2878
624	2	36-58:1952 189-210:874
627	4	25-48:2108 276-291:1253 334-351:1063 399-416:1680
628	4	22-37:2458 45-60:1250 82-98:1641 159-176:933
629	1	14-34:1660
630	1	12-38:1749
635	1	43-59:2213
636	1	13-34:2984
638	6	25-41:1898 103-119:1328 131-148:2506 180-203:1533 205-228:1303 245-260:1634
639	1	30-49:2416

Table 8
280

SEQ ID NO:	Number of Transmembrane Domains Predicted	Position of Transmembrane Region: TMPred Score
641	1	32-50:1597
642	1	284-299:1055
643	1	124-141:2071
645	1	92-108:1857
647	1	28-44:2543
649	2	43-58:1396 60-75:2059
650	3	5-35:1780 59-73:1361 80-103:1826
652	5	16-32:1576 72-87:1083 104-121:1825 145-160:1294 227-247:1337
654	1	39-53:1731
655	1	245-258:1771
656	1	58-81:2868
657	1	16-33:1894
658	1	290-310:2684
660	2	264-282:1757 383-403:1000
662	1	20-47:3001
663	2	18-33:892 108-126:1867
664	1	37-56:2054
665	1	369-387:2530
666	2	14-34:1939 187-208:1365
667	2	43-58:1060 155-170:2602
668	4	24-45:2509 98-119:2954 129-147:1343 183-201:2141
669	1	142-157:1775
670	1	33-49:2264
671	1	43-57:1794

Table 9
281

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.) *
1	337	673	874	788_13033
2	338			
3	339			
4	340	674	875	784_3746
5	341			
6	342	675	876	785_2855
7	343			
8	344			
9	345	676	877	785_1465
10	346	677	878	784_1644
11	347	678	879	784_4307
12	348			
13	349	679	880	787_1411
14	350	680	881	787_5936
15	351	681	882	784_4781
16	352	682	883	784_2486
17	353	683	884	790_28311
18	354	684	885	787_10206
19	355			
20	356			
21	357			
22	358	685	886	784_3665
23	359			
24	360	686	887	785_1105
25	361	687	888	787_7951
26	362	688	889	785_1538
27	363			
28	364	689	890	787_4539
29	365	690	891	790_26713
30	366	691	892	790_10585
31	367			
32	368	692	893	785_1092
33	369	693	894	784_5400
34	370			
35	371	694	895	790_17470
36	372			
37	373	695	896	784_844
38	374	696	897	787_9644
39	375	697	898	789_1867
40	376	698	899	785_612
41	377	699	900	785_1054
42	378	700	901	785_852
43	379	701	902	790_5231
44	380	702	903	784_5466
45	381			
46	382	703	904	790_21464
47	383	704	905	784_715
48	384	705	906	785_631
49	385	706	907	784_3853
50	386	707	908	790_10399

Table 9
282

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.) *
51	387	708	909	790_25607
52	388	709	910	790_10374
53	389	710	911	790_10504
54	390	711	912	790_21640
55	391	712	913	790_17957
56	392	713	914	787_71
57	393	714	915	791_1511
58	394	715	916	785_640
59	395	716	917	789_3732
60	396	717	918	787_5233
61	397	718	919	788_2575
62	398	719	920	790_4139
63	399	720	921	789_2499
64	400			
65	401	721	922	792_4675
66	402	722	923	784_2550
67	403	723	924	784_6192
68	404	724	925	787_7445
69	405			
70	406	725	926	787_5416
71	407	726	927	784_4167
72	408	727	928	784_5133
73	409			
74	410	728	929	784_10126
75	411			
76	412			
77	413	729	930	792_932
78	414	730	931	784_4665
79	415			
80	416			
81	417			
82	418	731	932	790_19568
83	419			
84	420			
85	421			
86	422			
87	423			
88	424			
89	425	732	933	784_1798
90	426			
91	427	733	934	790_1155
92	428	734	935	789_5186
93	429			
94	430			
95	431			
96	432			
97	433			
98	434			
99	435	735	936	790_8077
100	436	736	937	787_1058
101	437			

Table 9
283

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
102	438	737	938	784_929
103	439	738	939	788_10938
104	440			
105	441	739	940	787_5943
106	442	740	941	785_975
107	443	741	942	787_2691
108	444	742	943	785_3660
109	445	743	944	790_13070
110	446			
111	447	744	945	790_13664
112	448	745	946	790_24599
113	449			
114	450			
115	451			
116	452			
117	453	746	947	790_24595
118	454			
119	455	747	948	787_4919
120	456			
121	457			
122	458			
123	459	748	949	784_3534
124	460	749	950	784_4970
125	461			
126	462			
127	463	750	951	784_4845
128	464			
129	465	751	952	787_7638
130	466	752	953	785_1670
131	467	753	954	790_27718
132	468			
133	469	754	955	790_24877
134	470	755	956	790_9494
135	471	756	957	787_4525
136	472	757	958	784_939
137	473			
138	474			
139	475			
140	476			
141	477			
142	478	758	959	784_6707
143	479	759	960	788_11952
144	480	760	961	790_12052
145	481	761	962	790_3488
146	482	762	963	787_2489
147	483			
148	484			
149	485	763	964	792_3487
150	486			
151	487	764	965	785_395
152	488			

Table 9
284

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.) *
153	489			
154	490			
155	491			
156	492	765	966	785_3560
157	493			
158	494	766	967	785_1618
159	495			
160	496			
161	497			
162	498	767	968	787_4486
163	499	768	969	784_3498
164	500			
165	501			
166	502	769	970	784_5437
167	503	770	971	787_2054
168	504			
169	505	771	972	787_2155
170	506	772	973	790_15300
171	507			
172	508			
173	509			
174	510			
175	511	773	974	790_11357
176	512	774	975	789_2890
177	513			
178	514			
179	515			
180	516			
181	517	775	976	790_3760
182	518	776	977	784_4787
183	519			
184	520	777	978	787_4483
185	521			
186	522	778	979	785_598
187	523	779	980	790_2524
188	524	780	981	791_2994
189	525	781	982	784_4307
190	526			
191	527	782	983	790_11947
192	528			
193	529	783	984	787_6368
194	530	784	985	790_21374
195	531			
196	532	785	986	790_26925
197	533			
198	534			
199	535			
200	536	786	987	787_2905
201	537	787	988	784_5289
202	538	788	989	784_3437
203	539			

Table 9
285

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.) *
204	540			
205	541	789	990	785_158
206	542	790	991	784_1021
207	543	791	992	790_16269
208	544			
209	545			
210	546			
211	547	792	993	790_3621
212	548			
213	549	793	994	790_16011
214	550	794	995	790_18251
215	551	795	996	790_26204
216	552	796	997	790_17932
217	553	797	998	790_25384
218	554			
219	555	798	999	784_4771
220	556			
221	557	799	1000	784_9216
222	558	800	1001	787_7102
223	559	801	1002	784_8386
224	560	802	1003	790_21024
225	561			
226	562	803	1004	790_25301
227	563	804	1005	784_2437
228	564			
229	565	805	1006	784_3789
230	566	806	1007	787_4340
231	567	807	1008	788_8449
232	568	808	1009	790_17189
233	569	809	1010	790_3825
234	570	810	1011	784_7233
235	571			
236	572	811	1012	789_20
237	573			
238	574	812	1013	784_2129
239	575			
240	576	813	1014	787_5627
241	577	814	1015	787_7249
242	578			
243	579	815	1016	790_301
244	580	816	1017	784_1483
245	581	817	1018	784_5156
246	582	818	1019	787_2548
247	583			
248	584	819	1020	789_3213
249	585	820	1021	789_4901
250	586			
251	587	821	1022	790_24517
252	588			
253	589	822	1023	788_1187
254	590	823	1024	784_4265

Table 9
286

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
255	591	824	1025	784_603
256	592	825	1026	787_2104
257	593	826	1027	784_4819
258	594	827	1028	784_3677
259	595			
260	596			
261	597			
262	598			
263	599	828	1029	790_21539
264	600			
265	601			
266	602	829	1030	790_935
267	603			
268	604			
269	605	830	1031	787_3283
270	606	831	1032	787_7951
271	607	832	1033	790_13949
272	608	833	1034	784_2168
273	609	834	1035	785_1250
274	610	835	1036	784_9629
275	611			
276	612			
277	613			
278	614	836	1037	785_14
279	615	837	1038	790_24168
280	616	838	1039	787_4843
281	617			
282	618	839	1040	790_16366
283	619	840	1041	790_8044
284	620	841	1042	784_3590
285	621	842	1043	784_337
286	622	843	1044	785_706
287	623	844	1045	787_9834
288	624	845	1046	789_3409
289	625			
290	626	846	1047	787_3554
291	627	847	1048	790_8276
292	628	848	1049	785_3232
293	629	849	1050	784_3345
294	630	850	1051	790_18037
295	631			
296	632	851	1052	784_7084
297	633			
298	634			
299	635	852	1053	787_2278
300	636	853	1054	785_1867
301	637			
302	638	854	1055	787_2310
303	639	855	1056	784_2326
304	640			
305	641	856	1057	785_1538

Table 9
287

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No._SEQ ID NO.) *
306	642	857	1058	784_5007
307	643	858	1059	787_8999
308	644			
309	645			
310	646			
311	647	859	1060	787_5698
312	648	860	1061	790_29400
313	649			
314	650	861	1062	784_4813
315	651			
316	652	862	1063	784_9771
317	653	863	1064	790_10961
318	654	864	1065	790_11763
319	655	865	1066	784_5832
320	656			
321	657			
322	658	866	1067	790_16986
323	659	867	1068	785_3654
324	660	868	1069	785_102
325	661	869	1070	784_4307
326	662			
327	663			
328	664			
329	665			
330	666	870	1071	787_6896
331	667	871	1072	789_3174
332	668	872	1073	787_5591
333	669			
334	670			
335	671	873	1074	785_1003
336	672			

*784_XXX = SEQ ID NO: XXX of Attorney Docket No. 784, US Serial No. 09/488,725 filed 01/21/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

785_XXX = SEQ ID NO: XXX of Attorney Docket No. 785, US Serial No. 09/491,404 filed 01/25/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

787_XXX = SEQ ID NO: XXX of Attorney Docket No. 787, US Serial No. 09/496,914 filed 02/03/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

788_XXX = SEQ ID NO: XXX of Attorney Docket No. 788, US Serial No. 09/515,126 filed 02/28/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

Table 9
288

789_XXX = SEQ ID NO: XXX of Attorney Docket No. 789, US Serial No. 09/519,705 filed 03/07/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

790_XXX = SEQ ID NO: XXX of Attorney Docket No. 790, US Serial No. 09/540,217 filed 03/31/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

791_XXX = SEQ ID NO: XXX of Attorney Docket No. 791, US Serial No. 09/552,929 filed 04/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

792_XXX = SEQ ID NO: XXX of Attorney Docket No. 792, US Serial No. 09/577,408 filed 05/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference.

Table 10
289

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
1	337	10
2	338	100
3	339	101
4	340	102
5	341	103
6	342	104
7	343	105
8	344	106
9	345	107
10	346	108
11	347	109
12	348	11
13	349	110
14	350	111
15	351	112
16	352	113
17	353	114
18	354	115
19	355	116
20	356	117
21	357	118
22	358	119
23	359	12
24	360	120
25	361	121
26	362	122
27	363	123
28	364	124
29	365	125
30	366	126
31	367	127
32	368	128
33	369	129
34	370	13
35	371	130
36	372	131
37	373	132
38	374	133
39	375	134
40	376	135
41	377	136
42	378	137
43	379	138
44	380	139
45	381	14
46	382	140
47	383	141
48	384	142

Table 10
290

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
49	385	143
50	386	144
51	387	145
52	388	146
53	389	147
54	390	148
55	391	149
56	392	15
57	393	150
58	394	151
59	395	152
60	396	153
61	397	154
62	398	155
63	399	156
64	400	157
65	401	158
66	402	159
67	403	16
68	404	160
69	405	161
70	406	162
71	407	163
72	408	164
73	409	165
74	410	166
75	411	167
76	412	168
77	413	169
78	414	17
79	415	170
80	416	171
81	417	172
82	418	173
83	419	174
84	420	175
85	421	176
86	422	177
87	423	178
88	424	179
89	425	18
90	426	180
91	427	181
92	428	182
93	429	183
94	430	184
95	431	185
96	432	186
97	433	187

Table 10
291

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
98	434	188
99	435	189
100	436	19
101	437	190
102	438	191
103	439	192
104	440	193
105	441	194
106	442	195
107	443	196
108	444	197
109	445	198
110	446	199
111	447	2
112	448	20
113	449	200
114	450	201
115	451	202
116	452	203
117	453	204
118	454	205
119	455	206
120	456	207
121	457	208
122	458	209
123	459	21
124	460	210
125	461	211
126	462	212
127	463	213
128	464	214
129	465	215
130	466	216
131	467	217
132	468	218
133	469	219
134	470	22
135	471	220
136	472	221
137	473	222
138	474	223
139	475	224
140	476	225
141	477	226
142	478	227
143	479	228
144	480	229
145	481	23
146	482	230

Table 10
292

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
147	483	231
148	484	232
149	485	233
150	486	234
151	487	235
152	488	236
153	489	237
154	490	238
155	491	239
156	492	24
157	493	240
158	494	241
159	495	242
160	496	243
161	497	244
162	498	245
163	499	246
164	500	247
165	501	248
166	502	249
167	503	25
168	504	250
169	505	251
170	506	252
171	507	253
172	508	254
173	509	255
174	510	256
175	511	257
176	512	258
177	513	259
178	514	26
179	515	260
180	516	261
181	517	262
182	518	263
183	519	264
184	520	265
185	521	266
186	522	267
187	523	268
188	524	269
189	525	27
190	526	270
191	527	271
192	528	272
193	529	273
194	530	274
195	531	275

Table 10
293

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
196	532	276
197	533	277
198	534	278
199	535	279
200	536	28
201	537	280
202	538	281
203	539	282
204	540	283
205	541	284
206	542	285
207	543	286
208	544	287
209	545	288
210	546	289
211	547	29
212	548	290
213	549	291
214	550	292
215	551	293
216	552	294
217	553	295
218	554	296
219	555	297
220	556	298
221	557	299
222	558	3
223	559	30
224	560	300
225	561	301
226	562	302
227	563	303
228	564	304
229	565	305
230	566	306
231	567	307
232	568	308
233	569	309
234	570	31
235	571	310
236	572	311
237	573	312
238	574	313
239	575	314
240	576	315
241	577	316
242	578	317
243	579	318
244	580	319

Table 10
294

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
245	581	32
246	582	320
247	583	321
248	584	322
249	585	323
250	586	324
251	587	325
252	588	326
253	589	327
254	590	328
255	591	329
256	592	33
257	593	330
258	594	331
259	595	332
260	596	333
261	597	334
262	598	335
263	599	336
264	600	337
265	601	34
266	602	35
267	603	36
268	604	37
269	605	38
270	606	39
271	607	4
272	608	40
273	609	41
274	610	42
275	611	43
276	612	44
277	613	45
278	614	46
279	615	47
280	616	48
281	617	49
282	618	5
283	619	50
284	620	51
285	621	52
286	622	53
287	623	54
288	624	55
289	625	56
290	626	57
291	627	58
292	628	59
293	629	6

Table 10
295

SEQ ID NO of Full-length Nucleotide Sequence	SEQ ID NO of Full-length Peptide Sequence	SEQ ID NO in Priority Application USSN 60/311,261
294	630	60
295	631	61
296	632	62
297	633	63
298	634	64
299	635	65
300	636	66
301	637	67
302	638	68
303	639	69
304	640	7
305	641	70
306	642	71
307	643	72
308	644	73
309	645	74
310	646	75
311	647	76
312	648	77
313	649	78
314	650	79
315	651	8
316	652	80
317	653	81
318	654	82
319	655	83
320	656	84
321	657	85
322	658	86
323	659	87
324	660	88
325	661	89
326	662	9
327	663	90
328	664	91
329	665	92
330	666	93
331	667	94
332	668	95
333	669	96
334	670	97
335	671	98
336	672	99

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-336.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 99% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-336.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of any of the polynucleotides from SEQ ID NO: 1-336, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO: 337-672.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising of at least one of SEQ ID NO: 1-336.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.